

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 : Search time 4.92742 seconds
(without alignments)
1348.130 Million cell updates/sec

Title: US-09-941-947a-14
Perfect score: 806
Sequence: 1 MIRVCGYDVHRFNDGDHII.....LGFGRKGIAGVAVLIER 157

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCUTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575	71.3	164	US-09-252-991A-21228	Sequence 21228, A
2	496	61.5	165	US-09-328-352-7561	Sequence 7561, App
3	153	19.0	199	US-09-198-452A-585	Sequence 585, App
4	81	10.0	493	US-09-252-991A-24743	Sequence 24743, A
5	79.5	9.9	560	US-09-252-991A-26787	Sequence 26787, A
6	78.5	9.7	562	US-09-252-991A-32625	Sequence 32625, A
7	78	9.7	708	US-09-252-991A-32082	Sequence 32082, A
8	77	9.6	1477	US-09-206-942-71	Sequence 18149, A
9	75.5	9.4	220	US-09-252-991A-18149	Sequence 20039, A
10	75.5	9.4	400	US-09-252-991A-20039	Sequence 23689, A
11	73.5	9.1	898	US-09-252-991A-20237	Sequence 20237, A
12	73.5	9.1	1388	US-09-252-991A-20237	Sequence 3855, App
13	73	9.1	608	US-09-134-001C-3855	Sequence 24, App
14	73	9.1	1002	US-09-268-347-24	Sequence 28466, A
15	72.5	9.0	267	US-09-252-991A-28466	Sequence 21699, A
16	72	8.9	349	US-09-252-991A-21699	Sequence 22, App
17	71.5	8.9	1958	US-07-945-283-2	Sequence 1, App
18	70.5	8.7	272	US-09-071-035-22	Sequence 28696, A
19	70.5	8.7	317	US-08-973-275-1	Sequence 28696, A
20	70.5	8.7	411	US-09-252-991A-28863	Sequence 4306, App
21	70.5	8.7	1627	US-09-252-991A-28863	Sequence 6, App
22	70	8.7	156	US-09-134-001C-4306	Sequence 22184, A
23	70	8.7	203	US-08-549-680A-6	Sequence 314, App
24	70	8.7	203	US-08-549-680A-6	Sequence 9, App
25	70	8.7	250	US-09-252-991A-22184	
26	70	8.7	795	US-09-198-452A-314	
27	69.5	8.6	344	US-08-403-866-9	

28	69.5	8.6	489	US-09-252-991A-23781	Sequence 23781, A
29	69	8.6	441	US-09-252-991A-30150	Sequence 30150, A
30	69	8.6	549	US-08-354-240A-2	Sequence 2, App
31	69	8.6	550	US-08-354-240A-4	Sequence 4, App
32	69	8.6	550	US-08-354-240A-6	Sequence 23, App
33	69	8.6	550	US-08-867-352-23	Sequence 5, App
34	69	8.6	550	US-08-718-425-2	Sequence 2, App
35	69	8.6	550	US-08-718-425-5	Sequence 4, App
36	69	8.6	550	US-08-875-277A-2	Sequence 4, App
37	69	8.6	550	US-08-875-277A-4	Sequence 6, App
38	69	8.6	550	US-09-380-061B-6	Sequence 21, App
39	69	8.6	550	US-09-380-061B-11	Sequence 10, App
40	69	8.6	550	US-08-487-183A-20	Sequence 4, App
41	69	8.6	550	US-09-602-628-4	Sequence 6, App
42	69	8.6	550	US-09-602-628-6	Sequence 8, App
43	69	8.6	550	US-09-602-628-8	Sequence 12, App
44	69	8.6	550	US-09-602-628-12	Sequence 2, App
45	69	8.6	550	US-09-577-424-2	

ALIGNMENTS

```

RESULT 1
US-09-252-991A-21228
Sequence 21228, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21228
LENGTH: 164
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21228

Query Match      71.3%; Score 575; DB 4; Length 164;
Best Local Similarity 71.4%; Pred. No. 9.4e-63;
Matches 110; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

Cy      2  IRVWGVYDVHRFNDGDHIIIGGVKIPYKGLKMSDCGVVTHALADATGAAGDCKH 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      8  MRHGVYDVHRFNDGDHIIIGGVKIPYKGLKMSDCGVVTHALADATGAAGDCKH 67

Cy      62  FPDDEPFKAGSDVLLRHVGVYKGVKLVNADVTIIAQPGLPHVPGMRANIADL 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      68  FPDDEPFKAGSDVLLRHVGVYKGVKLVNADVTIIAQPGLPHVPGMRANIADL 127

Cy      122  RTVDVFNVAATTKKGFGRKGIAGVAVLI 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      128  GVAVDQVVAATTKKGFGRKGIAGVAVLI 161

RESULT 2
US-09-328-352-7561
Sequence 7561, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Brelton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC95-023A
CURRENT FILING DATE: 1999-06-04
CURRENT APPLICATION NUMBER: US/09/328,352
  
```

NUMBER OF SEQ ID NOS: 8252
 SEQ ID NO 7561
 LENGTH: 165
 TYPE: PRT
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-7561

Query Match 61.5%; Score 496; DB 4; Length 165;
 Best Local Similarity 61.0%; Pred. No. 4,9e-53;
 Matches 94; Conservative 25; Mismatches 35; Indels 0; Gaps 0;

QY 2 IRVGMGVDRFNDGHIILGKVIPEKGLFAHSDGVVLAALADALGADIGKH 61
 DB 10 IFIGQMDVHAEEGNEFVLAGVQPHHTGLKAHSDGVVLAALADALGADIGKH 69
 QY 62 PPDTPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRAIADL 121
 DB 70 PPDTPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRAIADL 129
 QY 122 ETDVDPEINVKATTTTEKGFEGKREGIAVQAVYLI 155
 DB 130 NVDLNDSIKATTTTEKGFEGKREGIAVQAVYLI 163

RESULT 3
 US-09-198-452A-585
 Sequence 585, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Gliteals, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 TITLE OF INVENTION: and treatment of infection
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 PRIOR FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 585
 LENGTH: 199
 TYPE: PRT
 ORGANISM: Chlamydia pneumoniae
 US-09-198-452A-585

Query Match 19.0%; Score 153; DB 4; Length 199;
 Best Local Similarity 29.4%; Pred. No. 9,3e-11;
 Matches 47; Conservative 26; Mismatches 75; Indels 12; Gaps 5;

QY 3 RVGMGVDRF---NDGHIILGKVIPEKGLFAHSDGVVLAALADALGADIGKH 56
 DB 40 RFGIGDSHREPLPESSTKCIIGITFDICPGFQANSDGDIIPHAICNAISSVTNKIILG 99
 QY 57 DIGKHPDTPNFKG-ADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRA 115
 DB 100 KVADELQF---RGITDSGIYIEALKSLK-PNKGISHVAITIGSRKFKLSALRQ 154
 QY 116 NTAADLETVDPIINVAATTTTEKGFEGKREGIAVQAVYLI 155
 DB 155 NTAQVNLTPDTDIGITATSGEGLSDFGCDGVCCFVLTIV 194

RESULT 4
 US-09-252-991A-24743
 Sequence 24743, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 24743
 LENGTH: 493
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-24743

Query Match 10.0%; Score 81; DB 4; Length 493;
 Best Local Similarity 30.9%; Pred. No. 0.26;
 Matches 29; Conservative 11; Mismatches 32; Indels 22; Gaps 5;

QY 5 GMGVDRFNDGHIILGKVIPEKGLFAHSDGVVLAALADALGADIGKH 62
 DB 185 GLGQVH---GDVVGGLALADHEDFRGTGDVADLAH-----VTLG--GGHV 229
 QY 63 PPDTPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRAIADL 96
 DB 230 ----DVAAGDDLPHPRHAFGAVGQGRHGLCPAD 258

RESULT 5
 US-09-252-991A-26787
 Sequence 26787, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 CURRENT FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 26787
 LENGTH: 560
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-26787

Query Match 9.9%; Score 79.5; DB 4; Length 560;
 Best Local Similarity 23.3%; Pred. No. 0.48;
 Matches 41; Conservative 34; Mismatches 78; Indels 23; Gaps 7;

QY 1 MIRVMGVDRFNDGHIILGKVIPEKGLFAHSDGVVLAALADALGADIGKH 44
 DB 81 LIRVGLAFRGHLRLADEAHLAALGLVSAVLLDLGVRGRLVHLLDQATVGDILQ 140
 QY 45 LA--DAIGAALADIGKHPDTPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIA 101
 DB 141 LGDDIVGTLA---VGHGFEDHLGDLAG-DGVVLDADQDAALGLGHRGLLDQAVIVE 196
 QY 102 QAPKMLPHVPGKRAIADLETVDPIINVAATTTTEKGFEGKREGIAVQAVYLI 157
 DB 197 QAAQLDHRHPRGRLGVAADALDLEVVGHLLAAGGQHAGVIGQAVLALAEAPALHR 252

RESULT 6
 US-09-252-991A-32625
 Sequence 32625, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18

PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 32625
 LENGTH: 562
 TYPE: PR1
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-32625

Query Match
 Best Local Similarity 24.3%; Score 78.5; DB 4; Length 562;
 Matches 35; Conservative 25; Mismatches 63; Indels 21; Gaps 5;

Db 16 GDIILGGVXIPY--EKGEASHSDGDVYVHIALADAILGAALGDIGKHPDIDPDKGAD 73
 Db 252 GGVVNDPQIRLQITGIGEAAGGGGLVQALFEGGRTTIGQ-----TORGAD 301
 Db 74 SRVLRHYVYGVKKGKYLVDVTIIAQAAPKMLPHV--GMRANIADLETDVDFINVA 132
 Db 302 ---LQHFROPFQGRQALLFAGAVY---QTAHRPIGRGDLADAGIGGFVDFPR 353
 Db 133 TTTEKGFEGRKEGIAVQAVLIE 156
 Db 354 LDORLHVHRQPAVAEHAVALLE 377

RESULT 7

US-09-252-991A-32082
 Sequence 32082, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 32082
 LENGTH: 708
 TYPE: PR1
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-32082

Query Match
 Best Local Similarity 27.8%; Score 78; DB 4; Length 708;
 Matches 42; Conservative 25; Mismatches 58; Indels 26; Gaps 10;

Db 21 LGGVXI-----PYEKGLE--ASHSDGDVYL--HALADAILGAALAD--IGKHPDIDPDK 70
 Db 206 VGVVYVVAAGPHATGIDPAHAHGVADVAGPEAGAVGVGGGGGFLVGGDSDBR 265
 Db 71 GADSRVLRHYVYGVK--EKGYKLVNADVTIIAQ--APKMLPHVPG--MRANIADLETDVD 126
 Db 266 ABD--FLTEHAHVAVPLEGG---RLDVVAAGQVAFELLDAAGDGLGFLGDIVEGSD 319
 Db 127 PINVXATTEKGFEGRKEGIAVQAVLIER 157
 Db 320 LVL-----LGGSLGADHGLGVQAVLADR 344

RESULT 8

US-09-206-942-71
 Sequence 71, Application US/09206942
 Patent No. 6432669

GENERAL INFORMATION:
 APPLICANT: Loosmore, Sheena M.
 APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michel H.
 TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 FILE REFERENCE: 1038-861 M15.1b
 CURRENT APPLICATION NUMBER: US/09/206,942
 CURRENT FILING DATE: 1998-12-08
 EARLIER APPLICATION NUMBER: 09/167,568
 EARLIER FILING DATE: 1998-10-07
 NUMBER OF SEQ ID NOS: 95
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 71
 LENGTH: 1477
 TYPE: PR1
 ORGANISM: Haemophilus influenzae
 US-09-206-942-71

Query Match
 Best Local Similarity 37.2%; Score 77; DB 4; Length 1477;
 Matches 29; Conservative 7; Mismatches 30; Indels 12; Gaps 3;

Db 29 EKGLEASHSDGDVYVHIALADAILGAALGDIGK--HPPDIDPDKGADSRVLRHYVYGVK 86
 Db 385 EKGRAIYVMDI---ALIDGINAAGSGDIATGTGFTSGHYLSIDN-----ALVK 434
 Db 87 EKGKLVNADVTIIAQAAP 104
 Db 435 TKEWLIDPDVDVIRAEEDP 452

RESULT 9

US-09-252-991A-18149
 Sequence 18149, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 18149
 LENGTH: 220
 TYPE: PR1
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-18149

Query Match
 Best Local Similarity 24.7%; Score 75.5; DB 4; Length 220;
 Matches 43; Conservative 25; Mismatches 67; Indels 39; Gaps 10;

Db 9 DVHRFNDGDHILGGVXIPYKGE-----LEASHDG-----DVVLA--IADAILGAA 54
 Db 46 DVLGGDDADHL---AIVPDHADTALLLEVHQGRGRARRRHRIRLHAGLAEIILGQA 102
 Db 55 LGDIGKHPDIDPDKGADSRVLRHYVYGVKKGKYLVDVTIIAQAAPKMLPHVPGAR 114
 Db 103 LGGAARDIAHVHDFDVLVDVAVVDRQA--GVA--GGAOLLDLDLEFFVVEVDAF----- 151
 Db 115 ANIAADLE--TDVDFINVKATTE-----KLGFEGRKRG--AVQAVVLIER 157
 Db 152 DVARDDHGVDRDLQVQVQAEQHVMPVYDLAGLGVVGRKGLFVAALAAVVLFR 205

RESULT 10

US-09-252-991A-20039
 Sequence 20039, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 20039
 LENGTH: 400
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-20039

Query Match 9.4%; Score 75.5; DB 4; Length 400;
 Best Local Similarity 23.8%; Pred. No. 0.89;
 Matches 44; Conservative 27; Mismatches 57; Indels 57; Gaps 9;

QY 2 IRVGMVDYHRFDDGHIILGKIPYKGLAH-----SDGVLALALD 47
 DB 121 VDVGVGGEAA--GQHPLIG--QHLLALQHRFRQTOPGLDAARLGGAVVEDTLD 173
 QY 48 AILGAALGDIKGFPTDPTNFKADSRVLLRHVGIK-----EKGVKLVNADY 97
 DB 174 PLIADLAVGVGQ-----DRVILQRDVLDVLTVMHPALDLFATGTALVHRM 221
 QY 98 -----TTIQAAPRLPHVGMKMANIAADLETVDYI--NVKATTE--KLGFEKX 144
 DB 222 ERNVWVVAIVAKORFEFGAIGHGTAAV--GPLQDLVHAIPHFDATGIGFAGFGRSGVE 280
 QY 145 EGIAY 149
 DB 281 DGIIV 285

RESULT 11
 US-09-252-991A-23689
 Sequence 23689, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 23689
 LENGTH: 898
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23689

Query Match 9.1%; Score 73.5; DB 4; Length 898;
 Best Local Similarity 22.7%; Pred. No. 5.3;
 Matches 44; Conservative 22; Mismatches 67; Indels 61; Gaps 8;

QY 4 VEMGIVY-----HRPVD-----GDHILGKIPYKGLA-----AHDGCVLHA 44
 DB 568 VGGDEARVYVMMRRDPRRAVGRGHIAAAGVLLVHRQVEVDFEHAESIAQAGLRP 627
 QY 45 LA-----DALGAALGDIKGFPTD-----PNEKADSRVLLRHVYG 83
 DB 628 LAQLAVERRRAADLDSAGHDALVMAAGDALIHHPDQAGAGTGGAGPLIV----- 682
 QY 84 IYKGVKLVNADY-----TTIQAAPRLPHVGMKMANIAADLETVDYVINVK 131

DB 683 -----LGHQLADPQAADAMVQLFRGLGEGQRAVLDNAVEAGGLVYDHTAAHRVTLA 738
 QY 132 ATTEKLGFEGRKE 145
 DB 739 AADLQAGGVEGTED 752

RESULT 12
 US-09-252-991A-20237
 Sequence 20237, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 20237
 LENGTH: 1388
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-20237

Query Match 9.1%; Score 73.5; DB 4; Length 1388;
 Best Local Similarity 24.2%; Pred. No. 10;
 Matches 36; Conservative 20; Mismatches 42; Indels 51; Gaps 7;

QY 17 DHILGKIPYKGL-----EAS--DGDVLA----- 43
 DB 273 DEDAFGRVAVVHGIFFLHYEALQIAGVLEVAHDHLDGDPVHQRTFGAGALTVDRS 332
 QY 44 --ALADAIIGAALGDIKGFPTDPTNFKADSRVLLRHVGIKGVKLVNADYITI 100
 DB 333 QAGLLAAGLASAMLAIEVEGEVAV-----AQLRRRVGVV--GIAVAVADVGLA 381
 QY 101 AARPMLPHV---PGMRANIAADLETVD 126
 DB 382 HQAPVGLHIAAPG-RAGVERADLAAYE 409

RESULT 13
 US-09-134-001C-3855
 Sequence 3855, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-607
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3855
 LENGTH: 608
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3855

Query Match 9.1%; Score 73; DB 4; Length 608;
 Best Local Similarity 21.7%; Pred. No. 3.4;
 Matches 28; Conservative 22; Mismatches 33; Indels 46; Gaps 6;

QY 18 HILIGVKIPYE-----KGLEHSDGVVLH-----ALADATGAAAG----- 56
 DB 372 HYMSGELVYDIDMTNKKGLFAGECDPFGHGNRLGANSLSATYCGYAGNAIKYE 431
 QY 57 DICKEHPDTPD-----NEKQADSRVLLRHVYGVKEKGYLVNADVT 98
 DB 432 NWKSTDTLDDSIPEARVKEKERPDHLIMRGTE-----NAVKLHREIG-EIMTANVT 484
 QY 99 IIAQAPKML 107
 DB 485 VVARENEKTL 493

RESULT 14
 US-09-268-347-24
 ; Sequence 24, Application US/09268347
 ; Patent No. 6335182

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

; FILE REFERENCE: 1038-860

; CURRENT APPLICATION NUMBER: US/09/268,347

; CURRENT FILING DATE: 1999-03-16

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 24

; LENGTH: 1002

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-268-347-24

Query Match 9.1%; Score 73; DB 4; Length 1002;

Best Local Similarity 24.8%; Pred. No. 7.3;

Matches 33; Conservative 16; Mismatches 46; Indels 38; Gaps 6;

QY 3 RYGMGYDVHFRNDSGDI---IILGVKIPYKGL-LEHSDGVVLHIALADATIGAAAGD 57

DB 489 KGVGVDGCGATGTHITDILVNSGDKVTLKAGDNLKVKQSGNFTYVLDDELGVKSV-- 546

QY 58 IGRFPDTPNFKGADSRVLLRHVYGVKEKGYLVNADVTIIAQAPKMLPHVPGKRAMI 117

DB 547 ---EFKDTENGANGAST-----KITKDGITI-----TPANDANG 577

QY 118 AADLETVDVPIVY 130

DB 578 AA--AIDADKIV 588

RESULT 15

US-09-252-991A-28466

; Sequence 28466, Application US/09252991A

; Patent No. 6531795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196-136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 28466

; LENGTH: 267

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-28466

Query Match 9.0%; Score 72.5; DB 4; Length 267;

Best Local Similarity 23.2%; Pred. No. 1.1;

Matches 35; Conservative 21; Mismatches 44; Indels 51; Gaps 8;

QY 41 VHALADAI-IGAAAL-----GDIQ-----KHPPD----- 64
 DB 54 VLRILAAVIGLACALAAABAEPLKIGVYVGVGDHGWTTQHEIGREBELVHFPDQVKT 113
 QY 65 --TDPNFKGADSRVLLRHV---YGIWEKGYLVNADVTIIAQAPKM-LPHVPGMR--A 115
 DB 114 SEVENVAEGADABRYIRHLAKDGYGLVFTTSFGVYNNPTAKVARQFPKVTFEHATGYKDR 173
 QY 116 NIADLETVDV-----FINVKATTEKLG 140
 DB 174 MGYLSRSRYEGRYVGGFLAAKMTSHKIGY 204

Search completed: January 29, 2004, 15:57:06
 Job time : 5.92742 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 10.7171 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947A-14

Perfect score: 806
Sequence: 1 MIRVGMYDHRNDSDHII.....LGFRRKRGIAVAVILIER 157

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubppaa/PC7_NEM_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubppaa/US06_NEM_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubppaa/US07_NEM_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubppaa/PC7US_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubppaa/US08_NEM_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubppaa/US09_NEM_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubppaa/US10_NEM_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubppaa/US60_NEM_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	806	100.0	157	10	US-09-934-903-10 Sequence 10, Appl
2	806	100.0	157	10	US-09-934-868-66 Sequence 66, Appl
3	806	100.0	157	11	US-09-941-947A-14 Sequence 14, Appl
4	496	61.5	170	15	US-10-174-410-12 Sequence 12, Appl
5	360.5	44.7	223	12	US-10-259-194A-162 Sequence 162, Appl
6	251	31.1	174	15	US-10-156-761-11502 Sequence 11502, A
7	247.5	30.7	158	12	US-10-128-713A-10 Sequence 10, Appl
8	213.5	26.5	160	10	US-09-738-626-6421 Sequence 6421, Ap
9	153	19.0	32	15	US-10-174-410-135 Sequence 135, Ap
10	153	19.0	199	12	US-10-289-762-585 Sequence 585, App
11	143	17.7	32	15	US-10-174-410-129 Sequence 129, App
12	138	17.1	32	15	US-10-174-410-133 Sequence 133, App
13	138	17.1	32	15	US-10-174-410-134 Sequence 134, App
14	138	17.1	32	15	US-10-174-410-136 Sequence 136, App
15	138	17.1	32	15	US-10-174-410-149 Sequence 149, App

16	133	16.5	32	15	US-10-174-410-125 Sequence 125, App
17	127	15.8	32	15	US-10-174-410-131 Sequence 131, App
18	126	15.6	32	15	US-10-174-410-127 Sequence 127, App
19	126	15.6	32	15	US-10-174-410-150 Sequence 150, App
20	125	15.5	32	15	US-10-174-410-128 Sequence 128, App
21	124	15.4	32	15	US-10-174-410-122 Sequence 122, App
22	124	15.4	32	15	US-10-174-410-140 Sequence 140, App
23	123	15.3	32	15	US-10-174-410-141 Sequence 141, App
24	123	15.3	32	15	US-10-174-410-124 Sequence 124, App
25	119	14.8	32	15	US-10-174-410-148 Sequence 148, App
26	116	14.4	32	15	US-10-174-410-132 Sequence 132, App
27	116	14.4	32	15	US-10-174-410-153 Sequence 153, App
28	115	14.3	32	15	US-10-174-410-130 Sequence 130, App
29	115	14.3	32	15	US-10-174-410-138 Sequence 138, App
30	115	14.3	32	15	US-10-174-410-134 Sequence 134, App
31	114	14.1	32	15	US-10-174-410-139 Sequence 139, App
32	114	14.1	32	15	US-10-174-410-142 Sequence 142, App
33	114	14.1	32	15	US-10-174-410-143 Sequence 143, App
34	114	14.1	32	15	US-10-174-410-156 Sequence 156, App
35	112	13.9	32	15	US-10-174-410-151 Sequence 151, App
36	112	13.9	32	15	US-10-174-410-223 Sequence 223, App
37	110	13.6	32	15	US-10-174-410-123 Sequence 123, App
38	109	13.5	32	15	US-10-174-410-158 Sequence 158, App
39	109	13.5	32	15	US-10-174-410-229 Sequence 229, App
40	109	13.5	32	15	US-10-174-410-230 Sequence 230, App
41	108	13.4	32	15	US-10-174-410-189 Sequence 189, App
42	106	13.2	32	15	US-10-174-410-126 Sequence 126, App
43	106	13.2	32	15	US-10-174-410-146 Sequence 146, App
44	106	13.2	32	15	US-10-174-410-155 Sequence 155, App
45	106	13.2	32	15	US-10-174-410-173 Sequence 173, App

ALIGNMENTS

RESULT 1
US-09-934-903-10
Sequence 10, Application US/0934903
Patent No. US20020102690A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odum, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. US20020102690A1ton, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rovielle, Pierre
APPLICANT: Piccaraggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OR INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: C1646 US NA
CURRENT APPLICATION NUMBER: US/09/934, 903
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229, 907
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 10
LENGTH: 157
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORFs
US-09-934-903-10
Query Match 100.0%; Score 806; DB 10; Length 157;
Best Local Similarity 100.0%; Pred. No. 3.5e-81;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRVGMYDHRNDSDHIIIGGVKIPYKGIHSHSDGVVJHALADATIGAAADIGK 60
DB 1 MIRVGMYDHRNDSDHIIIGGVKIPYKGIHSHSDGVVJHALADATIGAAADIGK 60
QY 61 HPPDTPNFKGADSRVILAHVYGIYKKGKYLWADVTIIAOPKMLPHYPKRARIAD 120

Db 61 HPPTDPEFKGADSRVLRHRYGIVKEKGYLVNADVTIIAQPRLPHYVGMGRANIAD 120
Qy 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157
Db 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157

RESULT 2

US-09-934-868-66
Sequence 66, Application US/09934868
Patent No. US20020137190A1
GENERAL INFORMATION:
APPLICANT: Kofas, Mattheos
APPLICANT: Odom, James M
APPLICANT: Schenck, Andreas J
TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN
FILE REFERENCE: CIL596 US NA
CURRENT APPLICATION NUMBER: US/09/934,868
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,858
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 81
SOFTWARE: Microsoft Office 97
SEQ ID NO 66
LENGTH: 157
TYPE: PRP
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ISPP
US-09-934-868-66

Query Match 100.0%; Score 806; DB 10; Length 157;
Best Local Similarity 100.0%; Pred. No. 3.5e-81;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRVGMGYDVHRRNDGDHIIIGGVKIPYKGLFAHSDGDVVLHALLADAIIGAAALGDIGK 60
Db 1 MIRVGMGYDVHRRNDGDHIIIGGVKIPYKGLFAHSDGDVVLHALLADAIIGAAALGDIGK 60
Qy 61 HPPTDPEFKGADSRVLRHRYGIVKEKGYLVNADVTIIAQPRLPHYVGMGRANIAD 120
Db 61 HPPTDPEFKGADSRVLRHRYGIVKEKGYLVNADVTIIAQPRLPHYVGMGRANIAD 120
Qy 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157
Db 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157

RESULT 3

US-09-941-947a-14
Sequence 14, Application US/09941947a
Publication No. US20030003528a1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia C.
APPLICANT: Cheng, Qiong
APPLICANT: DiCosimo, Deana J.
APPLICANT: Kofas, Mattheos
APPLICANT: Miller, Edward S. Jr.
APPLICANT: Odom, J. Martin
APPLICANT: Picataggio, Steve
APPLICANT: Rouviere, Pierre E.
TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
FILE REFERENCE: CIL903 US NA
CURRENT APPLICATION NUMBER: US/09/941,947a
CURRENT FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,858
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Microsoft Office 97
SEQ ID NO 14

LENGTH: 157
TYPE: PRP
ORGANISM: Methylobionas 16a
US-09-941-947a-14

Query Match 100.0%; Score 806; DB 11; Length 157;
Best Local Similarity 100.0%; Pred. No. 3.5e-81;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRVGMGYDVHRRNDGDHIIIGGVKIPYKGLFAHSDGDVVLHALLADAIIGAAALGDIGK 60
Db 1 MIRVGMGYDVHRRNDGDHIIIGGVKIPYKGLFAHSDGDVVLHALLADAIIGAAALGDIGK 60
Qy 61 HPPTDPEFKGADSRVLRHRYGIVKEKGYLVNADVTIIAQPRLPHYVGMGRANIAD 120
Db 61 HPPTDPEFKGADSRVLRHRYGIVKEKGYLVNADVTIIAQPRLPHYVGMGRANIAD 120
Qy 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157
Db 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157

RESULT 4

US-10-174-410-12
Sequence 12, Application US/10174410
Publication No. US20030073134A1
GENERAL INFORMATION:
APPLICANT: Louie, Gordon V.
APPLICANT: Gajiwala, Ketan S.
APPLICANT: Sauder, J. Michael
TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
FILE REFERENCE: 52498200300
CURRENT APPLICATION NUMBER: US/10/174,410
CURRENT FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: 60/259,058
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 336
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 170
TYPE: PRP
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: VARIANT
LOCATION: 1, 69, 108, 115
OTHER INFORMATION: Xaa = Selenomethionine
US-10-174-410-12

Query Match 61.5%; Score 496; DB 15; Length 170;
Best Local Similarity 61.8%; Pred. No. 7.7e-47;
Matches 97; Conservative 21; Mismatches 39; Indels 0; Gaps 0;

Qy 1 MIRVGMGYDVHRRNDGDHIIIGGVKIPYKGLFAHSDGDVVLHALLADAIIGAAALGDIGK 60
Db 3 LIRIGGFVHAHGBRPLIGVYPIHTGPIAHSDGVVLAHALLADAIIGAAALGDIGK 62
Qy 61 HPPTDPEFKGADSRVLRHRYGIVKEKGYLVNADVTIIAQPRLPHYVGMGRANIAD 120
Db 63 LPPTDIXQYNNADSRGLRBAFPQVQKGYKIGNVDTIIAQPRLPHYVGMGRANIAD 122
Qy 121 LETVDVFINVAKATTTEKLGFEGRKEGIAVOAVLIER 157
Db 123 LQCDIEQVNVLAITTEKLGFTGQDEGIACEAVALLIR 159

RESULT 5

US-10-259-194a-162
Sequence 162, Application US/10259194a
Publication No. US20040010815A1
GENERAL INFORMATION:
APPLICANT: Lange, Markus B.

```

RESULT 6
US-10-156-761-11502
Sequence 11502: Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCES: 249-262
CURRENT APPLICATION NUMBER: US/10-156-761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11502
LENGTH: 174
TYPE: RT
ORGANISM: Streptomyces avermitilis
US-10-156-761-11502

```

Query Match	30.7%	Score 247.5	DB 12	Length 158
Best Local Similarity	37.8%	Pred. No. 2.2e-19		
Matches 59	Conservative 24	Mismatches 66	Indels 7	Gaps 2
Qy	2	IRVMGDGYRFPMDGHIILGGVKIPPEKELIASHSDGDVHLALADAILGAALGDIKRI	61	
Db	1	MRVGLGTDVPIEIGRCRCMAAGILFERADCSGSHSDGVVHLCDDLSSAAGIGDLSV	60	
Qy	62	FPDIDPPEFKADSRVILRHVYGIYKEKGYLVNADVTIIAQAPPMLEPHVPGMRANIADL	121	
Db	61	EGTGREPMEDVSGARMIAEVRLLIENQOFVGNAAAVQVIGNRPT-----GPRDEAQKV	115	
Qy	122	ETDV--DFIVKATPTEKLGFEGRKEGIAVDAYVLI	155	
Db	116	LSDLIAPVVSATITDGLGELTGRGEGIAAMATNLV	151	

RESULT 8
US-09-738-626-6421
Sequence 6421, Application US/09738626
Publication NO. US20020197603A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, ACHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6421
LENGTH: 160
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6421

Query Match 26.5%; Score 213.5; DB 10; Length 160;
Best Local Similarity 32.9%; Pred. No. 1,3e-15;
Matches 51; Conservative 25; Mismatches 72; Indels 7; Gaps 2;

QY 3 RVGAGYVHRFNDGHIILGVKIPYKGLSHSDGVVLAALDAILGAALGIGKHF 62
DB 8 RVGIATDAHQIBAGKPCWIACLIFEGVDCGEGSDGVVAALIVDALISASGLGGSFV 67

QY 63 PPTDPFKAGDSRVLRLHYVGIYKKGKYLVDVTTIIAQAQKMLPHVPGKANTADLE 122
DB 66 GYGRPEYDVSQTOILKEVREILSAHGYVIGVMAQLVGTPEK-----GPRRERAGQVT 122

QY 123 TDV--DFINVKATTEKLGFBGRKEGIAVQAVLI 155
DB 123 SEIIGAPCCLSAITTDHMGFTGRSEGRASVATNAV 157

RESULT 9
US-10-174-410-135
Sequence 135, Application US/10174410
Publication No. US20030073134A1
GENERAL INFORMATION:

APPLICANT: Louie, Gordon V.
APPLICANT: Buchanan, Sean Grant
APPLICANT: Gajiwala, Ketan S.
APPLICANT: Sauder, J. Michael
TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
FILE REFERENCE: 52498200300
CURRENT APPLICATION NUMBER: US/10/174,410
CURRENT FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: 60/299,058
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 336
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 135
LENGTH: 32
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-10-174-410-135

Query Match 19.0%; Score 153; DB 15; Length 32;
Best Local Similarity 87.5%; Pred. No. 7e-10;
Matches 28; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 40 VVLAALDAILGAALGIGKHPPTDPNFKG 71
DB 1 VVLAALDAILGAALGIGKHPPTDPNFKG 32

RESULT 10
US-10-289-762-585
Sequence 585, Application US/10289762
Publication No. US2004006218A1
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/10/289,762

CURRENT FILING DATE: 2003-03-27
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 585
LENGTH: 199
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-10-289-762-585

Query Match 19.0%; Score 153; DB 12; Length 199;
Best Local Similarity 29.4%; Pred. No. 8.5e-09;
Matches 47; Conservative 26; Mismatches 75; Indels 12; Gaps 5;

QY 3 RVGAGYVHRF--NDGHIILGVKIPYKGLSHSDGVVLAALDAILGAALG 56
DB 40 RTGIGDSHRLPSSSTKPCILGSIYDPCGQANSDDGIIIFAIICNAISVTNKILG 99
QY 57 DIGHFPDTPNFKG-ADSRVLRHYVGIYKKGKYLVDVTTIIAQAQKMLPHVPGKRA 115
DB 100 KVADELQIOT---RGITSDGIYIEBALKSIR-PQKISHVAITEGSRPRFLCISALRQ 154

QY 116 NIADLETDVDFINVKATTEKLGFBGRKEGIAVQAVLI 155
DB 155 NIAQVMNLTPTDITGITAISGGLSDPCGSGVQCFCVLT 194

RESULT 11
US-10-174-410-129
Sequence 129, Application US/10174410
Publication No. US20030073134A1
GENERAL INFORMATION:

APPLICANT: Louie, Gordon V.
APPLICANT: Buchanan, Sean Grant
APPLICANT: Gajiwala, Ketan S.
APPLICANT: Sauder, J. Michael
TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
FILE REFERENCE: 52498200300
CURRENT APPLICATION NUMBER: US/10/174,410
CURRENT FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: 60/299,058
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 336
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 129
LENGTH: 32
TYPE: PRT
ORGANISM: C. perfringens
US-10-174-410-129

Query Match 17.7%; Score 143; DB 15; Length 32;
Best Local Similarity 78.1%; Pred. No. 8.9e-09;
Matches 25; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 40 VVLAALDAILGAALGIGKHPPTDPNFKG 71
DB 1 VVLAALDAILGAALGIGKHPPTDPNFKG 32

RESULT 12
US-10-174-410-133
Sequence 133, Application US/10174410
Publication No. US20030073134A1
GENERAL INFORMATION:
APPLICANT: Louie, Gordon V.
APPLICANT: Buchanan, Sean Grant
APPLICANT: Gajiwala, Ketan S.
APPLICANT: Sauder, J. Michael
TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
FILE REFERENCE: 52498200300
CURRENT APPLICATION NUMBER: US/10/174,410
CURRENT FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: 60/299,058

```

; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133
; LENGTH: 32
; TYPE: PRT
; ORGANISM: S. typhimurium
US-10-174-410-133

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALADAILGAAALGDIGKHPPTDPAFKG 71
1 VVHALTDALGAAALGDIGKLPPTDPAFKG 32

RESULT 13
US-10-174-410-134
; Sequence 134, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; FILE REFERENCE: 524982000300
; CURRENT FILING DATE: 2002-06-17
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 32
; TYPE: PRT
; ORGANISM: S. enterica
US-10-174-410-134

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALADAILGAAALGDIGKHPPTDPAFKG 71
1 VVHALTDALGAAALGDIGKLPPTDPAFKG 32

RESULT 14
US-10-174-410-136
; Sequence 136, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; FILE REFERENCE: 524982000300
; CURRENT FILING DATE: 2002-06-17
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Escherichia coli
```

```

US-10-174-410-136

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALADAILGAAALGDIGKHPPTDPAFKG 71
1 VVHALTDALGAAALGDIGKLPPTDPAFKG 32

RESULT 15
US-10-174-410-149
; Sequence 149, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; FILE REFERENCE: 524982000300
; CURRENT FILING DATE: 2002-06-17
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-174-410-149

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALADAILGAAALGDIGKHPPTDPAFKG 71
1 VVHALTDALGAAALGDIGKLPPTDPAFKG 32

Search completed: January 29, 2004, 16:21:17
Job time: 10.7171 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 4.61946 Seconds
(without alignments)
3268.453 Million cell updates/sec

Title: US-09-941-947a-14
Perfect score: 806
Sequence: 1 MIRWGAGYDVHRENDGDHIT.....LGFGRKREGIAVQAVVLI 157

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

1: PIR 76: *
2: PIR: *
3: PIR: *
4: PIR: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575	71.3	157	2 A83194	conserved hypotet
2	554	68.7	158	2 D82311	conserved hypotet
3	542	67.2	162	2 AB0408	2C-methyl-D-erythr
4	537	66.6	159	2 I55083	2C-methyl-D-erythr
5	537	66.6	159	2 A85924	2C-methyl-D-erythr
6	537	66.6	159	2 AD0856	2C-methyl-D-erythr
7	537	66.6	159	2 H91078	2C-methyl-D-erythr
8	534	66.3	158	2 F64156	hypothetical prote
9	520	64.5	157	2 D83663	gltx 5'-region con
10	520	64.5	157	2 D83663	hypothetical prote
11	505	62.7	160	2 G81073	YgbB/YacN family P
12	503	62.4	160	2 S70845	hypothetical prote
13	503	62.4	160	2 C81867	conserved hypotet
14	452	56.1	176	2 A82701	conserved hypotet
15	451	56.0	157	2 AE1104	B. subtilis YacN P
16	443	55.0	157	2 AE1104	hypothetical prote
17	425	52.7	379	2 S34980	hypothetical prote
18	424	52.6	155	2 C96953	hypothetical prote
19	414	51.4	380	2 A13359	2-C-methyl-D-eryth
20	408	50.6	382	2 F87464	conserved hypotet
21	403	50.0	161	2 E84978	hypothetical prote
22	403	50.0	165	2 AD2291	2-C-methyl-D-eryth
23	403	50.0	371	2 E81256	hypothetical prote
24	396	49.1	161	2 S73363	hypothetical prote
25	373	46.3	231	2 H96664	unknown protein, 3
26	372	46.2	400	2 AC7535	ispD/ispF bifuncti
27	372	46.2	400	2 AC7535	ISP/ISF bifuncti
28	345	42.8	409	2 G71936	hypothetical prote
29	343	42.6	406	2 D64647	conserved hypotet

30	329	40.8	399	2 G71314	conserved hypotet
31	306.5	38.0	165	2 P72351	conserved hypotet
32	301.5	37.4	156	2 F70382	conserved hypotet
33	289.5	35.9	161	2 F75542	conserved hypotet
34	285.5	35.4	158	2 B86949	hypothetical prote
35	276.5	34.3	159	2 C70607	hypothetical prote
36	269	33.4	240	2 B71615	YgbB protein, 1st
37	210	26.1	178	2 E81672	conserved hypotet
38	202	25.1	178	2 C71514	hypothetical prote
39	153	19.0	176	2 D72065	conserved hypotet
40	153	19.0	176	2 G86558	hypothetical prote
41	104.5	13.0	169	2 S73495	hypothetical prote
42	102.5	12.7	169	2 G64250	plasmid surface ex
43	83	10.3	593	2 T21510	hypothetical prote
44	81.5	10.1	294	2 C83497	ribosome ABC transp
45	81	10.0	434	2 T47545	monodehydroascorba

ALIGNMENTS

RESULT 1

A83194

conserved hypotet protein PA3627 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Dec-2002

C/Accession: A83194

R/Storer, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog

A/Reference number: A82950; UID:20437337; PMID:10984043

A/Accession: A83194

A/Status: preliminary

A/Molecule type: DNA

A/Restriction: 1-157 <STO>

A/Cross-references: GB:AE004782; GB:AE004091; MID:g9949772; PIDN:AA07015.1; GSPDB:GN001;

A/Experimental source: strain PA01

C/Genetics:

A/Geno: PA3627

C/Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match

Best Local Similarity 71.3%; Score 575; DB 2; Length 157;

Matches 110; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 2 IVWGAGYDVHRENDGDHIIIGSVKIPYKGLKSHSDGVVTLALADATIGAAAGDIGK 61

DB 1 MHWGAGYDVHRENDGDHIIIGSVKIPYKGLKSHSDGVVTLALADATIGAAAGDIGK 60

QY 62 FPDTPNFKGADSRVLRHRYGIVKKGKLVNADVTIIAQAQKXLPVFKGRANIADL 121

DB 61 FPDTPNFKGADSRVLRHRYGIVKKGKLVNADVTIIAQAQKXLPVFKGRANIADL 120

QY 122 ETDVDFINWAKATTEKELGFEGRKREGIAVQAVVLI 155

DB 121 GVAVDVQVNAKATTEKELGFEGRKREGIAVQAVVLI 154

RESULT 2

conserved hypotet protein VC0529 [imported] - Vibrio cholerae (strain N16961 serogr

C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Dec-2002

C/Accession: D82311

R/Heldberg, J.F.; Risen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dregoi, I.; Sellers, P.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; UID:20406833; PMID:10952301

A/Accession: D82311

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <HEI>
A:Cross-references: GB:AE004139; GB:AE003852; MID:9664953; PIDN:AAF93697.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype E1 Tor
A:Genetics:
A:Gene: VC0529
A:Map position: 1
C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 68.7%; Score 554; DB 2; Length 158;
Best Local Similarity 70.7%; Pred. No. 4.5e-44;
Matches 111; Conservative 16; Mismatches 30; Indels 0; Gaps 0;

QY 1 MIRVGKGVYVHRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHATADATLGAALGDIGK 60
DB 1 MIRGHPDVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHATDALLGAALGDIG 60
QY 61 HPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAAD 120
DB 61 HPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAAD 120
QY 121 LETVDVFINVKATTEKLGFEGRKEGIAVOAVVLI 157
DB 121 LETVDVFINVKATTEKLGFEGRKEGIAVOAVVLI 157

RESULT 3

AB0409
2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [imported] - Yersinia pestis (strain
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Dec-2002
C:Accession: AB0408
R:Parkhill, J.; Kren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Taranga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
H. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0408
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-162 <KUR>
A:Cross-references: GB:AL550842; PIDN:CAC92590.1; PID:G15981287; GSPDB:GN00175
C:Genetics:
A:Gene: 18pF
C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 67.2%; Score 542; DB 2; Length 162;
Best Local Similarity 68.6%; Pred. No. 5.9e-43;
Matches 107; Conservative 17; Mismatches 30; Indels 2; Gaps 1;

QY 2 IRVGMGVYVHRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHATADATLGAALGDIG 59
DB 1 MIRGHPDVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHATDALLGAALGDIG 60
QY 60 KHPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAA 119
DB 61 KHPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAA 120
QY 120 LETVDVFINVKATTEKLGFEGRKEGIAVOAVVLI 155
DB 121 LETVDVFINVKATTEKLGFEGRKEGIAVOAVVLI 155

RESULT 4

155083
2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [similarity] - Escherichia coli (st
C:Species: Escherichia coli
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 09-Dec-2002
C:Accession: I55083; F65055
R:Hi, C.; Ichikawa, J.K.; Ravetto, J.J.; Kuo, H.C.; Fu, J.C.; Clarke, S.
J. Bacteriol. 176, 6015-6022, 1994

A:Title: A new gene involved in stationary-phase survival located at 59 minutes on the
A:Reference number: I55083; MUID:95014035; PMID:7928962
A:Accession: I55083
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-159 <RES>
A:Cross-references: GB:I07942; MID:91036737; PIDN:AAA79837.1; PID:G433711
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F65055
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-159 <BLAT>
A:Cross-references: GB:AE00358; GB:U00096; MID:92367156; PIDN:AACT5788.1; PID:G1789103;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: y98b
C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 66.6%; Score 537; DB 2; Length 159;
Best Local Similarity 69.5%; Pred. No. 1.7e-42;
Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVGMGVYVHRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHATADATLGAALGDIGK 61
DB 1 MIRGHPDVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHATDALLGAALGDIGK 60
QY 62 HPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAAD 121
DB 61 HPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAAD 120
QY 122 ETVDDVFINVKATTEKLGFEGRKEGIAVOAVVLI 155
DB 121 GCHMDVAVKATTEKLGFEGRKEGIAVOAVVLI 154

RESULT 5

AB5924
2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [similarity] - Escherichia coli (st
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Dec-2002
C:Accession: AB5924
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: AB5924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <STO>
A:Cross-references: GB:AE005174; MID:912517200; PIDN:AAG57853.1; GSPDB:GN00145; UNKGP:24C
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: y98b
C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 66.6%; Score 537; DB 2; Length 159;
Best Local Similarity 69.5%; Pred. No. 1.7e-42;
Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVGMGVYVHRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHATADATLGAALGDIGK 61
DB 1 MIRGHPDVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHATDALLGAALGDIGK 60
QY 62 HPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAAD 121
DB 61 HPPTDPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAARKLPHVPGKRAMIAAD 120
QY 122 ETVDDVFINVKATTEKLGFEGRKEGIAVOAVVLI 155

A/Cross-references: GB:Z99104; GB:AL0009126; NID:G2632267; PIDN:CA81867.1; PID:G2632358
 A/Experimental source: strain 168
 R/Ogasawara, N.; Nakai, H.
 DNA Res. 1, 1-14, 1994
 A/Title: Systematic sequencing of the 180 kbpase region of the *Bacillus subtilis* chrom
 A/Reference number: S63567; MUID:96051385; PMID:7584024
 A/Accession: S66120
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-158 <CGA>
 A/Cross-references: EMBL:U26185; NID:G467326; PIDN:BA05325.1; PID:G467479
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
 R/Gagnon, Y.; Breton, R.; Futerer, H.; Pelchat, M.; Grunberg-Manago, M.; Lapointe, J.
 J. Biol. Chem. 269, 7473-7482, 1994
 A/Title: Clustering and co-transcription of the *Bacillus subtilis* genes encoding the am
 A/Reference number: A53402; MUID:9411772; PMID:7510287
 A/Accession: A53402
 A/Molecule type: DNA
 A/Residues: 137-158 <GAG>
 A/Cross-references: GB:U14580
 A/Genetics:
 A/Experimental source: strain C-125
 A/Status: YACN
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 Query Match 66.3%; Score 534; DB 2; Length 158;
 Best Local Similarity 65.6%; Pred. No. 3.2e-42;
 Matches 103; Conservative 18; Mismatches 36; Indels 0; Gaps 0;

Query 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 Db 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 Qy 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 Db 1 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 1 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 Qy 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 157
 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 157
 Db 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 157
 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 157

RESULT 10
 D83663
 hypotetical protein BH0108 (imported) - *Bacillus halodurans* (strain C-125)
 C/Species: *Bacillus halodurans*
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Dec-2002
 C/Accession: D83663
 R/Nakami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
 Nucleic Acids Res. 28, 4317-4331, 2000
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A/Reference number: A83650; MUID:20512582; PMID:11058132
 A/Accession: D83663
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-157 <STO>
 A/Cross-references: GB:AP001507; GB:BA000004; NID:G10172612; PIDN:BA803827.1; GSPDB:GN00
 A/Experimental source: strain C-125
 A/Status: preliminary
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 C/Supfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 A/Status: preliminary
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 Query Match 64.5%; Score 520; DB 2; Length 157;
 Best Local Similarity 66.5%; Pred. No. 6.2e-41;
 Matches 103; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

Query 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 Db 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 Qy 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 Db 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120

Query 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 155
 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 155
 Db 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 155
 121 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 155

RESULT 11
 G81073
 ygdB/Yacn family protein NMB512 (imported) - *Neisseria meningitidis* (strain MCS8 serogr
 C/Species: *Neisseria meningitidis*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Dec-2002
 C/Accession: G81073
 R/Fretellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.B.; Eisen, J.A
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 Li, H.; Qin, H.; Yamahavean, J.; Gill, J.; Scarlato, V.; Mastignani, V.; Pizzia, M.
 Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ve
 A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MCS8.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: G81073
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-160 <TEF>
 A/Cross-references: GB:AE002501; GB:AE002098; NID:G7226755; PIDN:AAF41868.1; PID:G722675
 A/Experimental source: serogroup B, strain MCS8
 A/Genetics:
 A/Status: NMB512
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 C/Supfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 Query Match 62.7%; Score 505; DB 2; Length 160;
 Best Local Similarity 60.9%; Pred. No. 1.5e-39;
 Matches 95; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

Query 2 IRVGGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 61
 2 IRVGGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 61
 Db 4 IRVGGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 63
 4 IRVGGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 63

Query 62 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 121
 62 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 121
 Db 64 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 123
 64 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 123

Query 122 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 157
 122 LETVDVFINVKATTTKTLGFEGRKEGIAVQAVLIER 157
 Db 124 GIDISCVNKKKTENKTLGFEGRKEGIAVQAVLIER 159
 124 GIDISCVNKKKTENKTLGFEGRKEGIAVQAVLIER 159

RESULT 12
 S70845
 hypotetical protein 1 - *Haemophilus ducreyi*
 C/Species: *Haemophilus ducreyi*
 C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 09-Dec-2002
 C/Accession: S70845
 R/Palmer, K.L.; Munson Jr., R.S.
 Mol. Microbiol. 18, 821-830, 1995
 A/Title: Cloning and characterization of the genes encoding the haemolysin of *Haemophilu*
 A/Reference number: S70843; MUID:96422469; PMID:885086
 A/Accession: S70845
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-160 <PAL>
 A/Cross-references: EMBL:U23175; NID:G151070; PIDN:AAC43539.1; PID:G151073
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1995
 C/Supfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 A/Status: preliminary
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
 Query Match 62.4%; Score 503; DB 2; Length 160;
 Best Local Similarity 63.1%; Pred. No. 2.4e-39;
 Matches 99; Conservative 17; Mismatches 41; Indels 0; Gaps 0;

Query 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 Db 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 1 MIRVGAGYVHRFNDGDHIIIGGVKIPYKGLKSHSDGDVLAHALADALIGAAALGDIK 60
 Qy 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120
 61 HEPDTPNPKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120

Db 61 LFPDTPQYNNISKLLIYAVQVCKGQGINIDITIIACQPKRPHIDNRQIAND 120
 QY 121 LETDVDFINVAATTTTEKLGFEGRKEGIAVAVLIER 157
 121 LNCBIDQINKATTTTEKLGFTGREGIACEAVALLER 157

RESULT 13

Conserved hypothetical protein NMA1712 [imported] - Neisseria meningitidis (strain Z2491)
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Dec-2002
 C:Accession: C81867
 R:Parikh, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, R.; Parkhill, J.; Jorgensen, K.; Leather, S.; Moule, S.; Mangall, K.; Quail, M.A.; Rajandream, A.; Holt, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mangall, K.; Quail, M.A.; Rajandream, A.; Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: AB1775; MUID:2022556; PMID:10761919
 A:Accession: C81867
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <PAR>
 A:Cross-references: GB:AL162756; GB:AL157959; NID:q7380091; PION:CA84940.1; PID:q738035
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA1712
 C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 62.4%; Score 503; DB 2; Length 160;
 Best Local Similarity 60.9%; Pred. No. 2,4e-39;
 Matches 95; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

QY 2 IRVGMGVDFRFDGHIILGKVIYKGLASHSQGVVLAHLADIIIGAAALGDIGK 61
 Db 4 IIRGGGVDFHQLTEKGLIGVBIPEKGLCHSDADILHAYVTLALGAGLGDIGSH 63
 QY 62 FPPDTPNFKGADSRVLLRHVYGVYKKEKGYLVNADVTIIIAQAPKMLPHVPGKRIANIADL 121
 Db 64 FPPDTPNFKGADSRVLLRHVYGVYKKEKGYLVNADVTIIIAQAPKMLPHVPGKRIANIADL 123

QY 122 ETDVDPIVNAATTTTEKLGFEGRKEGIAVAVLIER 157
 Db 124 GIDISCNTIKGKTNEKLGIDYRMEGIESQAAVLVR 159

RESULT 14

Conserved hypothetical protein XPI294 [imported] - Xylella fastidiosa (strain 945c)

C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Dec-2002
 C:Accession: AB2701
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: AB2715; MUID:20355717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: AB2701
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-176 <SIM>
 A:Cross-references: GB:AE003962; GB:AE003849; NID:g9106270; PION:AA64103.1; GSPDB:GN001
 A:Experimental source: strain 945c

A:Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Ferra, S.C.; Franco, M.C.; Frohm, J.D.; Tuncel, M.L.; Kemper, B.L.; Kitzajima, J.P.; Klegler, J.E.; Kurnace, E.E.; Laigz, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tsubako, M.H.; Vallada, H.; Van Sluys, W.A.; Vezjowski-Almeida, S.; Vettore, A.L.; Z

Query Match 56.1%; Score 452; DB 2; Length 176;
 Best Local Similarity 53.8%; Pred. No. 1.4e-34;
 Matches 84; Conservative 31; Mismatches 41; Indels 0; Gaps 0;

QY 2 IRVGMGVDFRFDGHIILGKVIYKGLASHSQGVVLAHLADIIIGAAALGDIGK 61
 Db 15 IIRGGGVDFHQLTEKGLIGVBIPEKGLCHSDADILHAYVTLALGAGLGDIGSH 74
 QY 62 FPPDTPNFKGADSRVLLRHVYGVYKKEKGYLVNADVTIIIAQAPKMLPHVPGKRIANIADL 121
 Db 75 FPPDTPNFKGADSRVLLRHVYGVYKKEKGYLVNADVTIIIAQAPKMLPHVPGKRIANIADL 134

QY 122 ETDVDPIVNAATTTTEKLGFEGRKEGIAVAVLIER 157
 Db 135 EYQPCVSVKATTSKGLPFGKRGIAVAVLIER 170

RESULT 15

B. subtilis Yach protein homolog lmo0236 [imported] - Listeria monocytogenes (strain EGD-AE1104)
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Dec-2002
 C:Accession: AE1104
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Kert, U.

Science 294, 849-852, 2001
 A:Authors: Kretz, U.; Kuhn, M.; Kunst, F.; Knapkat, G.; Madueno, E.; Maitournam, A.; Matok, C.; Schuster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AE1104
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-157 <GLA>
 A:Cross-references: GB:NC 003210; PION:CAD00763.1; PID:g16409601; GSPDB:GN00177
 A:Experimental source: strain EGD-A

C:Genetics:
 A:Gene: lmo0236
 C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 56.0%; Score 451; DB 2; Length 157;
 Best Local Similarity 58.6%; Pred. No. 1.5e-34;
 Matches 92; Conservative 19; Mismatches 46; Indels 0; Gaps 0;

QY 1 MIRVGMGVDFRFDGHIILGKVIYKGLASHSQGVVLAHLADIIIGAAALGDIGK 60
 Db 1 IIRGGGVDFHQLTEKGLIGVBIPEKGLCHSDADILHAYVTLALGAGLGDIGSH 60
 QY 61 FPPDTPNFKGADSRVLLRHVYGVYKKEKGYLVNADVTIIIAQAPKMLPHVPGKRIANIADL 120
 Db 61 FPPDTPNFKGADSRVLLRHVYGVYKKEKGYLVNADVTIIIAQAPKMLPHVPGKRIANIADL 120
 QY 121 LETDVDFINVAATTTTEKLGFEGRKEGIAVAVLIER 157
 Db 121 LHAQSAQVNAATTTTEKLGFTGREGIASIAVALLER 157

Search completed: January 29, 2004, 15:55:32
 Job time : 5.61946 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 2.95645 Seconds
(without alignments)
2497.314 Million cell updates/sec

Title: US-09-941-947A-14

Perfect score: 806

Sequence: 1 MTRVGMGYDVHFRNDGDHII.....LGFEGKXGGLAVQAVLLER 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	575	71.3	157	1	ISPF_PSEAE
2	554	68.7	158	1	ISPF_VIBCH
3	551	68.4	158	1	ISPF_VIBVU
4	550	68.2	158	1	ISPF_VIBB2
5	542	67.2	162	1	ISPF_YR2PE
6	537	66.6	159	1	ISPF_ECOL6
7	537	66.6	159	1	ISPF_ECOL1
8	537	66.6	159	1	ISPF_SALTI
9	537	66.6	159	1	ISPF_SALTY
10	534	66.3	158	1	ISPF_BACSU
11	534	66.3	158	1	ISPF_HAEIN
12	529	65.6	158	1	ISPF_PASMD
13	520	64.5	157	1	ISPF_BACHD
14	505	62.7	160	1	ISPF_NEIMB
15	503	62.4	160	1	ISPF_HAEIM
16	503	62.4	160	1	ISPF_NEIMA
17	498	61.8	156	1	ISPF_CLOPE
18	498	61.8	156	1	ISPF_RALSO
19	470	58.3	163	1	ISPF_XANAC
20	456	56.6	166	1	ISPF_XANCP
21	452	56.1	176	1	ISPF_XYLF4
22	451	56.0	157	1	ISPF_LISMO
23	446	55.3	160	1	ISPF_FUSNN
24	443	55.0	157	1	ISPF_LISIN
25	429.5	53.3	160	1	ISPF_THETN
26	425	52.7	179	1	ISPF_CHOAB
27	424	52.6	155	1	ISPF_RHOCA
28	415	51.5	434	1	ISPF_RHIME
29	414	51.4	390	1	ISPF_BRIME
30	414	51.4	407	1	ISPF_RHIL0
31	408	50.6	382	1	ISPF_CAUCR
32	403	50.0	161	1	ISPF_BUCAT
33	403	50.0	165	1	ISPF_ANASP

ALIGNMENTS

RESULT	ID	ISPF_PSEAE	STANDARD	PRT	157 AA
34	34	403	50.0	167	1
35	35	403	50.0	371	1
36	36	396	49.1	161	1
37	37	383	47.5	236	1
38	38	373	46.3	231	1
39	39	372	46.2	400	1
40	40	345	42.8	409	1
41	41	343	42.6	406	1
42	42	329.5	40.9	387	1
43	43	329.5	40.9	157	1
44	44	329	40.8	399	1
45	45	306.5	38.0	165	1

RESULT	ID	ISPF_PSEAE	STANDARD	PRT	157 AA
AC	AC	P57708; O9HY03;			
DT	DT	16-OCT-2001 (Rel. 40, last sequence update)			
DT	DT	16-OCT-2001 (Rel. 40, last sequence update)			
DE	DE	28-FEB-2003 (Rel. 41, last annotation update)			
DE	DE	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)			
DE	DE	(MECPs) (MECPD-synthase).			
OS	OS	ISPF OR PA3627.			
OC	OC	Pseudomonas aeruginosa.			
OC	OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
OC	OC	Pseudomonadaceae; Pseudomonas.			
OX	OX	NCBI_taxid=287;			
RN	RN	(1)			
RP	RP	SEQUENCE FROM N.A.			
RC	RC	STRAIN=ATCC 15692 / PA01;			
RX	RX	MEDLINE=20437337; PubMed=10984043;			
RA	RA	Stover C.K., Phan X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	RA	Hickey M.V., Brinkman P.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,			
RA	RA	Garber R.L., Goltzer S.N., Tolentino E., Westbrock-Medman S., Yuan Y.,			
RA	RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Lashby K., Lim R.M.,			
RA	RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	RA	Feiler J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;			
RT	RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an			
RT	RT	opportunistic pathogen."			
RU	RU	Nature 406:959-964(2000).			
CC	CC	- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-			
CC	CC	PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND			
CC	CC	CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO			
CC	CC	2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).			
CC	CC	- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-			
CC	CC	methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate			
CC	CC	+ CMP.			
CC	CC	- PATHWAY: Nonaevonolone terpenoid biosynthesis pathway; fifth step.			
CC	CC	- SIMILARITY: BELONGS TO THE ISPF FAMILY.			
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	CC	use by non-profit institutions as long as its content is in no way			
CC	CC	modified and this statement is not removed. Usage by and for commercial			
CC	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	CC	or send an email to license@isb-sib.ch).			
DR	DR	EMBL; AE004782; AA067015.1;			
DR	DR	PIR; A83194; A83194.			
DR	DR	HAWAP; WF_001077; -; 1.			
DR	DR	InterPro; IPR003526; YgBB.			
DR	DR	Pfam; PF02542; YgBB; 1.			
DR	DR	TIGRFAMs; TIGR0151; ISPF; 1.			
DR	DR	PROSITE; PS01350; ISPF; 1.			
KW	KW	lysase; isoprene biosynthesis; Complete proteome.			
SQ	SQ	SEQUENCE 157 AA; 16662 MW; C39C377B6403BCA CRC64;			

Query Match

71.3%; Score 575; DB 1; Length 157;

Best Local Similarity 71.4%; Pred. No. 8,3e-47;
Matches 110; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 2 IRVGMGVYDVRHENDGHIILIGVYKIPYKGLKHAHSDGVYLAHALADAILGAAALGIGK 61
DB 1 MIRIGHGVYDVRHFGEGEPIITLGGVRIPIKGLVAHSDGVYLAHALADAILGAAALGIGK 60
QY 62 FPDTPDFPKGADSRVLLRHYGYIVKRGYLVNADVTIIQAAPKMLPHVFGKRAITADL 121
DB 61 FPDTPDFPKGADSRVLLRHYGYIVKRGYLVNADVTIIQAAPKMLPHVFGKRAITADL 120
QY 122 ETDVDFINVKATTTKXGFGKRGKGIAGVAVLIER 155
DB 121 GVAVDGVNVAATTTKXGFGKRGKGIAGVAVLIER 154

RESULT 2

ISPF_VIBCH STANDARD; PRT; 158 AA.

ID ISPF_VIBCH STANDARD; PRT; 158 AA.
AC 09KDU1;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
DE (MECPS) (MECPS-synthase).
GN ISPF OR V05529.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EI for N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Tsayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin J., Richardson D.,
RA Ermolaeva K.D., Vamathevan J., Bases S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae".
RL Nature 406:477-483(2000).
CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
CC PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
CC CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
CC 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
CC + CMP.
CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.
CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: AB004139; AAF93697.1;
CC PIR: D82311; D82311.
CC TIGR: V00529;
CC HAMAP: MF_00107; -; 1.
CC InterPro: IPR003526; YgDB.
CC Pfam: PF02542; YgDB; 1.
CC TIGRFAMs: TIGR00151; ispf; 1.
CC PROSITE: PS01350; ISPF; 1.
CC Lyase: Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 158 AA; 16837 MW; 42B0AA40FFBFC6A CRC64;

Query Match 58.7%; Score 554; DB 1; Length 158;
Best Local Similarity 70.7%; Pred. No. 7.6e-45;
Matches 111; Conservative 16; Mismatches 30; Indels 0; Gaps 0;

QY 1 MIRVGMGVYDVRHENDGHIILIGVYKIPYKGLKHAHSDGVYLAHALADAILGAAALGIGK 60
DB 1 MIRIGHGVYDVRHFGEGEPIITLGGVRIPIKGLVAHSDGVYLAHALADAILGAAALGIGK 60
QY 61 FPDTPDFPKGADSRVLLRHYGYIVKRGYLVNADVTIIQAAPKMLPHVFGKRAITADL 120
DB 61 FPDTPDFPKGADSRVLLRHYGYIVKRGYLVNADVTIIQAAPKMLPHVFGKRAITADL 120
QY 121 LETDVGFINVKATTTKXGFGKRGKGIAGVAVLIER 157
DB 121 LETDVGFINVKATTTKXGFGKRGKGIAGVAVLIER 157

RESULT 3

ISPF_VIBVU STANDARD; PRT; 158 AA.

ID ISPF_VIBVU STANDARD; PRT; 158 AA.
AC 08DC59;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
DE (MECPS) (MECPS-synthase).
GN ISPF OR VY11583.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWCP6;
RX Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CWCP6";
RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-
CC phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and
CC CMP. Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into
CC 2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (By similarity).
CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
CC + CMP.
CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.
CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: AB016802; AA010006.1;
CC HAMAP: MF_00107; -; 1.
CC InterPro: IPR003526; YgDB.
CC Pfam: PF02542; YgDB; 1.
CC TIGRFAMs: TIGR00151; ispf; 1.
CC PROSITE: PS01350; ISPF; 1.
CC Lyase: Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 158 AA; 17084 MW; 0E888F8050160112 CRC64;

Query Match 68.4%; Score 551; DB 1; Length 158;
Best Local Similarity 69.0%; Pred. No. 1.5e-44;
Matches 107; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MIRVGMGVYDVRHENDGHIILIGVYKIPYKGLKHAHSDGVYLAHALADAILGAAALGIGK 60
DB 1 MIRIGHGVYDVRHFGEGEPIITLGGVRIPIKGLVAHSDGVYLAHALADAILGAAALGIGK 60

QY 61 HFPDTPNFKGADSRVLLRHVYGIKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120
 DB 61 HFPDTPNFKGADSRVLLRHVYGIKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120
 OY 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVVLI 155
 DB 121 LETDIRNINVAATTEKLGFEGRKEGIAVQAVVLI 155

RESULT 4

ISPF_VIBRA STANDARD; PRT; 158 AA.
 ID ISPF_VIBRA
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECP-synthase).
 GN ISPF OR VP2558.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 CC NCBI_TaxId=670;
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nishina M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
 RT Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 distinct from that of V. cholerae.";
 RL Lancel 361:743-749(2003).
 CC -1- FUNCTION: Converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and
 CMP. Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into
 2C-methyl-D-erythritol 3,4-cyclodiphosphate and CMP (by similarity).
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC -1- PATHWAY: Nonaevalonate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP005081; BAC60821.1; -
 DR HAMAP; MF 00107; -; 1.
 DR PROSITE; PS01350; ISPF, 1.
 KM Isoprene biosynthesis; lyase; Complete proteome.
 SQ SEQUENCE 158 AA; 16949 MW; F8C9FPAE03D3D1D CRC64;

Query Match 68.2%; Score 550; DB 1; Length 158;
 Best local similarity 69.0%; Pred. No. 1; 8e-44;

Matches 107; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MIRVGAGVDFHRRDNDGHIILGVKIPYKGLFAHSDGVDVLIHALADALIGAAALGDICK 60
 DB 1 MIRVGAGVDFHRRDNDGHIILGVKIPYKGLFAHSDGVDVLIHALADALIGAAALGDICK 60
 OY 1 MIRVGAGVDFHRRDNDGHIILGVKIPYKGLFAHSDGVDVLIHALADALIGAAALGDICK 60
 DB 1 MIRVGAGVDFHRRDNDGHIILGVKIPYKGLFAHSDGVDVLIHALADALIGAAALGDICK 60
 QY 61 HFPDTPNFKGADSRVLLRHVYGIKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120
 DB 61 HFPDTPNFKGADSRVLLRHVYGIKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120
 OY 61 HFPDTPNFKGADSRVLLRHVYGIKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120
 DB 61 HFPDTPNFKGADSRVLLRHVYGIKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120
 QY 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVVLI 155
 DB 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVVLI 155
 OY 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVVLI 155
 DB 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVVLI 155

RESULT 5
 ISPF_YERPE STANDARD; PRT; 162 AA.
 ID ISPF_YERPE
 AC Q6ZBP7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECP-synthase).
 GN ISPF OR YPO3360 OR Y0829.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 CC NCBI_TaxId=332;
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,
 RT Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
 RA Perry R.D.,
 RT Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -1- FUNCTION: Converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and
 CMP. Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into
 2C-methyl-D-erythritol 3,4-cyclodiphosphate and CMP (by similarity).
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC -1- PATHWAY: Nonaevalonate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ414156; CAC92590.1; -
 DR EMBL; AE013685; AAM84414.1; ALT_INIT.
 DR PIR; AB0408; AB0408.
 DR HAMAP; MF 00107; -; 1.
 DR InterPro; IPR003526; YGPB.
 DR Pfam; PF0242; YGPB; 1.
 DR TIGRPFAM; TIGR00151; ISPF, 1.
 DR PROSITE; PS01350; ISPF, 1.
 KM Lyase; isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 162 AA; 17182 MW; 9A295411D3DF1EB CRC64;

Query Match 67.2%; Score 542; DB 1; Length 162;
 Best local similarity 68.6%; Pred. No. 1e-43;

Matches 107; Conservative 17; Mismatches 30; Indels 2; Gaps 1;

OY 2 IRVAGVGVDFHRRF--NDGHIILGVKIPYKGLFAHSDGVDVLIHALADALIGAAALGDIG 59

Db 1 MRIGGDFVHKGAGNSGSPILIGVRI:PYEKGGLAHSDGDVALHAADALGAAALGDIG 60
 Qy 60 KHPEDTPNPNFGASRVLLRHVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPCGRANTIA 119
 Db 61 KLPFDTPNPAFAGASRGRLRRAVYRIIAKGYKGLNLTITIAQAPKPAHPHPCGRVNLAE 120
 Qy 120 DLRTDVPFNKATTTTEKLGPEGRKEGIAVOAVL 155
 Db 121 DLQCHMDINVKATTTTQLGFTGEGIGIACRAVALL 156

RESULT 6

ISPF_ECOL6
 ID ISPF_ECOL6 STANDARD; PRT; 159 AA.

AC Q8REJ6;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECDP-synthase).
 ISPF OR C313.
 OS Escherichia coli 06
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NX NCBI_TaxID=217992;
 RN [1]

SEQUENCE FROM N.A.

RC STRAIN=06:H1 / ATCC 700928;
 RX MEDLINE=2338234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 Raeko D., Buckles E.L., Lau S.-R., Boutin A., Hackett J., Stroud D.,
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

CC -1- FUNCTION: Converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and
 CMP. Also converts 4-diphosphocytidyl-2C-methyl-D-erythritol into
 2C-methyl-D-erythritol 3,4-cyclodiphosphate and CMP (by similarity).
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP

CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.ch).

CC EMBL: AE016765; AAN01762.1; -

DR HAMAP: MF 00107; -

DR InterPro: IPR003526; YgbB.

DR Pfam: PF02542; YgbB; 1.

DR TIGRFAMs: TIGR00151; ISPF; 1.

DR PROSITE: PS01350; ISPF; 1.

DR Isoprene biosynthesis; lyase; Complete proteome.

KN SEQUENCE 159 AA; 16867 MM; SP04573623BA4549 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;

Best Local Similarity 69.5%; Pred. No. 3e-43;

Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

Qy 2 IRVAGGVDFHFNQGDHIIIGVKKIPYKGLAEHSDGDVALHAADALGAAALGDIG 61

Db 1 MRIGGDFVHKGAGNSGSPILIGVRI:PYEKGGLAHSDGDVALHAADALGAAALGDIG 60

Qy 62 FPDTPNPAFAGASRGRLRRAVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPCGRANTIA 121

Db 61 FPDTPNPAFAGASRGRLRRAVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPCGRANTIA 120
 Qy 122 FPDTPNPAFAGASRGRLRRAVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPCGRANTIA 155
 Db 121 GCHMDVNVKATTTTEKLGFTGEGIGIACRAVALL 154

RESULT 7

ISPF_ECOL1
 ID ISPF_ECOL1 STANDARD; PRT; 159 AA.

AC P36663;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECDP-synthase).
 ISPF OR MECS OR E2746 OR Z4054 OR EC63600 OR SF2769.
 OS Escherichia coli
 OS Escherichia coli 0157:H7, and
 OS Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NX NCBI_TaxID=562, 83334, 623;
 RN [1]

SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=MP180;
 RX MEDLINE=95014035; PubMed=7928962;

RA U.C., Ichikawa J.K., Ravetto J.J., Kuo H.-C., Fu J.C., Clarke S.;
 RT "A new gene involved in stationary-phase survival located at 59
 minutes on the Escherichia coli chromosome";
 RL J. Bacteriol. 176:6015-6022(1994).

RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.

RC SPECIES=E.coli;
 RX MEDLINE=20183914; PubMed=10694574;

RA Herz S., Wungstentweck J., Schuhr C.A., Hecht S., Luettgen H.,
 Sagner S., Fellemeier W., Hisenreich W., Zenk M.H., Bacher A.,
 RA Rohdich F.;
 RT "Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-
 2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-
 cyclodiphosphate";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2486-2490(2000).

RN [3] SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC SPECIES=E.coli;
 RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Mayhew G.F.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).

RP SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Joshi G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck S.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Annatharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;

KN "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";

RT Nature 409:529-533(2001).

RA [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Kihara S., Shida T., Hattori M., Sasaki K., Ogasawara N., Yasunaga T.,
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=2272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Sun J., Yang F., Zhang X., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.,
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
 through comparison with genomes of *Escherichia coli* K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [8]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RC SPECIES=E.coli;
 RX MEDLINE=99420866; PubMed=10493123;
 RA Pountoulakis M., Takacs W.-F., Berndt P., Langen H., Takacs B.,
 RT "Enrichment of low abundance proteins of *Escherichia coli* by
 hydrophobic chromatography.";
 RL Electrophoresis 20:2181-2185(1999).
 RN [9]
 RP FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP.
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC COFACTOR: MAGNESIUM OR MANGANESE.
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L07942; AAA79837.1;
 DR EMBL: A230738; AA644656.1;
 DR EMBL: AB038256; BAA95145.1;
 DR EMBL: U29579; AA69256.1;
 DR EMBL: AE000358; AAC75788.1;
 DR EMBL: AE005502; AAG57853.1;
 DR EMBL: AP002562; BAB37023.1;
 DR EMBL: AE015291; AAN44258.1;
 DR PIR: A85924; A85924.
 DR PIR: H91078; H91078.
 DR PIR: I55083; I55083.
 DR PDB: 1GK1; 16-MAY-02.
 DR PDB: 1Y8; 30-JUN-02.
 DR PDB: 1KXJ; 18-JUN-02.
 DR PDB: 1KXK; 18-JUN-02.
 DR Ecocore: EGI1816; ISPF.
 DR HAMAP: MF_00107; -; 1.
 DR InterPro: IPR003526; Ygdb.
 DR Pfam: PF02542; Ygdb.1.
 DR TIGRFAMs: TIGR00151; ispf.1.
 DR PROSITE: PS01350; ISPF.1.
 KM Lyase; isoprene biosynthesis; Magnesium; Manganese; Complete proteome;
 3D-structure.

SQ SEQUENCE 159 AA; 16897 MW; 9FC556323A6293 CRC64;
 Query Match 66.6%; Score 537; DB 1; Length 159;
 Best Local Similarity 69.5%; Pred. No. 36-43;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;
 QY 2 IIVKMGVYDHRFNDGHIILGCVKIPYEKGLFAHSDGVDVLAALDAIAGAALDIGN 61
 DB 1 MGIQGFVDVHAGSGGPIIGVRIPEYKGLAHSDGVVLAHALDALLGAALDIGN 60
 QY 62 PFDTPNFGADSRVLRHVIGIVKEKGYLVNADVTIIAQPXKLPHVPGKANIADL 121
 DB 61 PFDTPNFGADSRVLRHVIGIVKEKGYLVNADVTIIAQPXKLPHVPGKANIADL 120
 QY 122 EVDVDFINKATTTKLGPEKKGIAVAVVLI 155
 DB 121 GAMDVNVKATTTKLGFTGRGSLGACVALL 154
 RESULT 8
 ISPF_SALTI
 ID ISPF_SALTI STANDARD; PRT; 159 AA.
 AC 082472;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECS) (MECD-synthase).
 GN ISPF OR STY3054 OR 12830.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 ON NCBI TaxID=601;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21534947; PubMed=11677608;
 RA Parthali V., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsis K.,
 RA Krogh A., Larsen T.S., Leach S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burdard V., Kodymiani V., Schwartz D.C., Blatner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RT and CT18.";
 RL Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: Converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and
 CMP. Also converts 4-diphosphocytidyl-2C-methyl-D-erythritol into
 2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (by similarity).
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch.

CC EMBL: AL627276; CAD06035.1; -
 DR HAMBL: ABO16843; AAO70387.1; -
 DR HAMBL: ME_00107; - 1
 DR InterPro: IPR003526; Ygdb.
 DR Pfam: PF02542; Ygdb; 1.
 DR TIGRFAMs: TIGR00151; ISPF; 1.
 DR PROSITE: PS01350; ISPF; 1.
 KM Lyase; Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 159 AA; 15885 MW; 9FD9263625662939 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;
 Best Local Similarity 69.5%; Pred. No. 3e-43;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVGMGVVHREDFNDGHIILGKVKIPYKGLKSHSDGCVLHATADATIGAAAGDICK 61
 DB 1 MRIGGFVHAFGEGPPIIGVRIPEYKGLKSHSDGCVLHATADATIGAAAGDICK 60
 QY 62 PPDTDPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAAPKMLPHVPGKRAIAADL 121
 DB 61 PPDTDPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAAPKMLPHVPGKRAIAADL 120
 QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVYLI 155
 DB 121 GCHMDVNVKATTEKLGFEGRKEGIAVQAVYLI 154

RESULT 9

ID ISPF SALTY STANDARD; PRT; 159 AA.

AC Q6ZME7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECPD-synthase).
 GN ISPF OR STM2929.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SSGC1412 / AFCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sandereson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney B.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2."

RT Nature 413:852-856(2001).
 CC -1- FUNCTION: Converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and
 CMP. Also converts 4-diphosphocytidyl-2C-methyl-D-erythritol into
 2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (By similarity).
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BESONONS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AB008833; AAL21809.1; -

DR STYGENE: SG7272?; ISPF.
 DR HAMBL: ME_00107; - 1
 DR InterPro: IPR003526; Ygdb.
 DR Pfam: PF02542; Ygdb; 1.
 DR TIGRFAMs: TIGR00151; ISPF; 1.
 DR PROSITE: PS01350; ISPF; 1.
 KM Lyase; Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 159 AA; 16899 MW; 9FD8BD2C95666939 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;
 Best Local Similarity 69.5%; Pred. No. 3e-43;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVGMGVVHREDFNDGHIILGKVKIPYKGLKSHSDGCVLHATADATIGAAAGDICK 61
 DB 1 MRIGGFVHAFGEGPPIIGVRIPEYKGLKSHSDGCVLHATADATIGAAAGDICK 60
 QY 62 PPDTDPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAAPKMLPHVPGKRAIAADL 121
 DB 61 PPDTDPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAAPKMLPHVPGKRAIAADL 120
 QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVYLI 155
 DB 121 GCHMDVNVKATTEKLGFEGRKEGIAVQAVYLI 154

RESULT 10

ID ISPF BACSU STANDARD; PRT; 158 AA.

AC Q06756;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECPD-synthase).
 GN ISPF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94171772; PubMed=7510287;
 RX Gagnon Y., Breton R., Putzer H., Pelchat M., Grunberg-Manago M.,
 RA Lapointe J.;
 RT "Clustering and co-transcription of the *Bacillus subtilis* genes
 RT encoding the aminoacyl-tRNA synthetases specific for glutamate and for
 RT cysteine and the first enzyme for cysteine biosynthesis.";
 RL J. Biol. Chem. 269:7473-7482(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT "Systematic sequencing of the 180 kilobase region of the *Bacillus*
 RT subtilis chromosome containing the replication origin.";
 RL DNA Res. 1:1-14(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunze F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bortero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortis R., Boursier L., Brans A., Braun M., Brigell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano Y., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Eutlian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holleppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kaashara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kunita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudaga B., Park S.H.,
 RA Paro V., Pohl T.M., Portetalle D., Portolillo S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield S.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanaka S., Tanaka T., Terstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Vairi A., Wambuit R., Wedler E., Wedler H., Weitzengger T.,
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein E.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*,"
 RT Nature 350:249-256(1997).
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL 2-
 CC PHOSPHATE INTO 2-C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CC CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL INTO
 CC 2-C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 CC + CMP.
 CC -1- PATHWAY: Nonaevonate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, L4589; AAA21795.1; --
 CC EMBL, D6185; BAA05325.1; --
 CC EMBL, Z99104; CAB11867.1; --
 CC PIR, F69741; F69741.
 CC Subtilist; BG10153; ispf.
 CC HAMAP; MF 00107; -- 1.
 CC InterPro; IPR003526; Ygbb.
 CC Pfam; PF02542; Ygbb; 1.
 CC TIGRFAMs; TIGR00151; ispf; 1.
 CC PROSITE; PS01350; ispf; 1.
 CC Lyase; Isoprene biosynthesis; Complete proteome.
 CC SEQUENCE 158 AA; 17126 MW; ED11D03EC3752BD0 CRC64;
 SO
 Query Match 66.3%; Score 534; DB 1; Length 158;
 Best Local Similarity 65.6%; Pred. No. 5, 6e-43;
 Matches 103; Conservative 18; Mismatches 36; Indels 0; Gaps 0;
 QY 1 MIRVGAGYDVRHNDGHIILIGVKIPYKGLAHSDDGVYLAALADALIGAAAGDIDK 60
 DB 1 MRLIGGFVHGVHVRPLILIGIEIPIYKGLGSHDADVLHTAVLADLGAAGVGGDIDK 60
 QY 61 HPPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAQKMLPHVPGKMANIAAD 120
 DB 61 HPPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAQKMLPHVPGKMANIAAD 120
 QY 121 LETVDVFINVAKATTTKLGFEKRGKGIYAVOAVLTIR 157
 DB 121 LEADVSQVAVKATTTKLGFEKRGKGIYAVOAVLTIR 157
 RESULT 11
 ISPF HAIRIN STANDARD; PRT; 158 AA.
 AC P44815;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DB 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MCCS) (MCCD-synthase).
 GN ISPF OR H10671.
 OS Haemophilus influenzae.
 OC Pasteurellales; Pasteurellales; Haemophilus.
 OC NCBI_TaxId=727;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=9550630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M.,
 RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uiterback T.R., Hanna M.C., Nguyen D.T., Staudt D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 RA *rd*,"
 RA Science 269:496-512(1995).
 RM [2]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE=98384070; PubMed=9719565;
 RA Fountoulakis M., Duranville J.F., Roder D., Evers S., Berndt P.,
 RA Langen H.;
 RT "Reference map of the low molecular mass proteins of *Haemophilus*
 RT *influenzae*,"
 RT Electrophoresis 19:1819-1827(1998).
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL 2-
 CC PHOSPHATE INTO 2-C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CC CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL INTO
 CC 2-C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 CC + CMP.
 CC -1- PATHWAY: Nonaevonate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U32750; AAC22331.1; --
 CC PIR, F64156; P64156.
 CC PDB; 1UN1; 21-AUG-02.
 CC TIGR; H10671; --
 CC HAMAP; MF 00107; -- 1.
 CC InterPro; IPR003526; Ygbb.
 CC Pfam; PF02542; Ygbb; 1.
 CC TIGRFAMs; TIGR00151; ispf; 1.
 CC PROSITE; PS01350; ispf; 1.
 CC Lyase; Isoprene biosynthesis; Complete proteome; 3D-structure.
 CC SEQUENCE 158 AA; 17194 MW; DC34BR347DEC2BF2 CRC64;
 SO
 Query Match 66.3%; Score 534; DB 1; Length 158;
 Best Local Similarity 65.6%; Pred. No. 5, 6e-43;
 Matches 103; Conservative 20; Mismatches 34; Indels 0; Gaps 0;
 QY 1 MIRVGAGYDVRHNDGHIILIGVKIPYKGLAHSDDGVYLAALADALIGAAAGDIDK 60
 DB 1 MRLIGGFVHGVHVRPLILIGIEIPIYKGLGSHDADVLHTAVLADLGAAGVGGDIDK 60
 QY 61 HPPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAQKMLPHVPGKMANIAAD 120
 DB 61 HPPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIAQAQKMLPHVPGKMANIAAD 120

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 157
 DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 157

RESULT 12

ISPF_PASMTU STANDARD; PRT; 158 AA.

ID ISPF_PASMTU

AC 157954;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)

DE (MECPS) (MECDP-synthase).

GN ISPF OR PM1609.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Pasteurella.

OX NCBI_TaxID=747;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PM70;

RA MEDLINE=2145866; PubMed=11248100;

RA May B., Zhang Q., Li L., Paustian M.L., Whitam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70."

RT "Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001)."

CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.

CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fifth step.

CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: AB006198; AA03693.1; -

DR HAMAP: MF_00107; -; 1.

DR InterPro: IPR003526; YqdB.

DR Pfam: PF02542; YqdB; 1.

DR TIGRFAMs: TIGR00151; ispf; 1.

DR PROSITE: PS01350; ISPF; 1.

KW Lyase; Isoprene biosynthesis; Complete proteome.

SQ SEQUENCE 158 AA; 16555 MW; 3589CBA59147171E CRC64;

Query Match 65.6%; Score 529; DB 1; Length 158;

Best Local Similarity 67.7%; Pred. No. 1.7e-42;

Matches 105; Conservative 15; Mismatches 35; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHRENDGHIILGKVIPEYKGLBAHSDGVDVLAHALADALGAALGDIK 60

DB 1 MIRVGKGVYHRENDGHIILGKVIPEYKGLBAHSDGVDVLAHALADALGAALGDIK 60

QY 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLVAADVTIIAQAQKLPHVPGKRAITAA 120

DB 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLVAADVTIIAQAQKLPHVPGKRAITAA 120

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

ID ISPF_BACHD STANDARD; PRT; 157 AA.

AC 09KGF7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)

DE (MECPS) (MECDP-synthase).

GN ISPF OR BH0108.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

halodurans and genomic sequence comparison with Bacillus subtilis."

RT Nucleic Acids Res. 28:4317-4331(2000).

CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.

CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fifth step.

CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: AP001507; BAB03827.1; -

DR HAMAP: MF_00107; -; 1.

DR InterPro: IPR003526; YqdB.

DR Pfam: PF02542; YqdB; 1.

DR TIGRFAMs: TIGR00151; ispf; 1.

DR PROSITE: PS01350; ISPF; 1.

KW Lyase; Isoprene biosynthesis; Complete proteome.

SQ SEQUENCE 157 AA; 16751 MW; 0205862C7B59BA99 CRC64;

Query Match 64.5%; Score 520; DB 1; Length 157;

Best Local Similarity 66.5%; Pred. No. 1.1e-41;

Matches 103; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHRENDGHIILGKVIPEYKGLBAHSDGVDVLAHALADALGAALGDIK 60

DB 1 MIRVGKGVYHRENDGHIILGKVIPEYKGLBAHSDGVDVLAHALADALGAALGDIK 60

QY 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLVAADVTIIAQAQKLPHVPGKRAITAA 120

DB 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLVAADVTIIAQAQKLPHVPGKRAITAA 120

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

ID ISPF_BACHD STANDARD; PRT; 157 AA.

AC 09KGF7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)

DE (MECPS) (MECDP-synthase).

GN ISPF OR BH0108.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

halodurans and genomic sequence comparison with Bacillus subtilis."

RT Nucleic Acids Res. 28:4317-4331(2000).

CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.

CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fifth step.

CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: AP001507; BAB03827.1; -

DR HAMAP: MF_00107; -; 1.

DR InterPro: IPR003526; YqdB.

DR Pfam: PF02542; YqdB; 1.

DR TIGRFAMs: TIGR00151; ispf; 1.

DR PROSITE: PS01350; ISPF; 1.

KW Lyase; Isoprene biosynthesis; Complete proteome.

SQ SEQUENCE 157 AA; 16751 MW; 0205862C7B59BA99 CRC64;

Query Match 64.5%; Score 520; DB 1; Length 157;

Best Local Similarity 66.5%; Pred. No. 1.1e-41;

Matches 103; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHRENDGHIILGKVIPEYKGLBAHSDGVDVLAHALADALGAALGDIK 60

DB 1 MIRVGKGVYHRENDGHIILGKVIPEYKGLBAHSDGVDVLAHALADALGAALGDIK 60

QY 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLVAADVTIIAQAQKLPHVPGKRAITAA 120

DB 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLVAADVTIIAQAQKLPHVPGKRAITAA 120

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

QY 121 LETDNDPINKATTTETKLGFEGRKEGIAVOAVLIER 155

DB 121 LQCDIEQVNVKATTTETKLGFTGRGEGIAVALLIR 155

ID ISPF_BACHD STANDARD; PRT; 157 AA.

AC 09KGF7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)

DE (MECPS) (MECDP-synthase).

GN ISPF OR BH0108.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

halodurans and genomic sequence comparison with Bacillus subtilis."

RT Nucleic Acids Res. 28:4317-4331(2000).

CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECDP-synthase).
 GN ISPF OR NM31512.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tetselin H., Saunders N.C., Heidelberg J., Jeffries A.C., Nelson K.B.,
 Eisen J.A., Ketchum K.A., Hood D.W., Heden J.F., Dodson R.J.,
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecko A., Parksey D.S., Blair M., Citron H., Clark E.B.,
 Cotton M.D., Uetack T.R., Khouf H., Qin H., Vamathevan J.,
 Gill J., Scarlato V., Maignani V., Pizza M., Grand G., Sun J.,
 Salath H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.";
 RL Science 287:1809-1815 (2000).
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; AE002501; AAF41868.1; -
 DR PIR; G81073; G81073.
 DR TIGR; NMB1512; -
 DR HAMAP; MF_00107; -; 1.
 DR InterPro; IPR003526; Ygdb.
 DR Pfam; PF02342; Ygdb; 1.
 DR TIGRFAMs; TIGR00151; ISPF; 1.
 DR PROSITE; PS01350; ISPF; 1.
 KW Lyase; Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 160 AA; 17019 MW; 9C223BCED4D9BF3D CRC64;
 Query Match 62.7%; Score 505; DB 1; Length 160;
 Best Local Similarity 60.9%; Pred. No. 2.9e-40;
 Matches 95; Conservative 21; Mismatches 40; Indels 0; Gaps 0;
 QY 2 IRVKGVDVRFMDGHIILGGVPIPEKGLIAHSDGDDVTLALADLIGAAAGDGGK 61
 DB 4 IRVKGVDVRFMDGHIILGGVPIPEKGLIAHSDGDDVTLALADLIGAAAGDGGK 63
 QY 62 FPDIDPFKAGDSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVPGMRANIADL 121
 DB 64 FPDIDPFKAGDSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVPGMRANIADL 123
 QY 122 ETDVDFINVAATTEKLGFBGREGIAVQAVTLIER 157
 DB 124 GIDISCVNIGKTKNEKDGFGREGIEAQAVALIVR 159

DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)
 DE (MECPs) (MECDP-synthase).
 GN ISPF.
 OS Haemophilus ducreyi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 NCBI_TaxID=730;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=35000;
 RX MEDLINE=96422469; PubMed=8825086;
 RA Palmer K.P., Munson R.S. Jr.;
 RT "Cloning and characterization of the genes encoding the hemolysin of
 Haemophilus ducreyi.";
 RL Mol. Microbiol. 18:871-830 (1995).
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
 + CMP.
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; U32175; AAC33539.1; -
 DR PIR; S70845; S70845.
 DR HAMAP; MF_00107; -; 1.
 DR InterPro; IPR003526; Ygdb.
 DR Pfam; PF02542; Ygdb; 1.
 DR TIGRFAMs; TIGR00151; ISPF; 1.
 DR PROSITE; PS01350; ISPF; 1.
 KW Lyase; Isoprene biosynthesis.
 SQ SEQUENCE 160 AA; 17437 MW; D7F1F2F518F7000E CRC64;
 Query Match 62.4%; Score 503; DB 1; Length 160;
 Best Local Similarity 63.1%; Pred. No. 4.5e-40;
 Matches 99; Conservative 17; Mismatches 41; Indels 0; Gaps 0;
 QY 1 MTRVKGVDVRFMDGHIILGGVPIPEKGLIAHSDGDDVTLALADLIGAAAGDGGK 60
 DB 1 MTRVKGVDVRFMDGHIILGGVPIPEKGLIAHSDGDDVTLALADLIGAAAGDGGK 60
 QY 61 HPPDIDPFKAGDSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVPGMRANIADL 120
 DB 61 LPPDIDPFKAGDSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVPGMRANIADL 120
 QY 121 LETVDVFINVAATTEKLGFBGREGIAVQAVTLIER 157
 DB 121 LNCDDIDQINIVATTEKLGFBGREGIAVQAVTLIER 157

Search completed: January 29, 2004, 15:50:41
 Job time : 3.95645 secs

RESULT 15
 ISPF HAEDU
 ID ISPF HAEDU STANDARD; PRT; 160 AA.
 AC 047956;
 DT 30-MAY-2000 (Rel. 39, Created)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 11.7026 Seconds
(without alignments)
3461.979 Million cell updates/sec

Title: US-09-941-947a-14

Perfect score: 806
Sequence: 1 MIRVGMGYDVHRENDGDHIL.....LAFEGRGKGIAYQAVLIER 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 segs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL.23.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	569	70.6	159	16	Q8EBR3
2	551	68.4	158	16	Q8DCS9
3	537	66.6	159	16	Q8FEJ6
4	453	56.2	157	16	Q8KCS5
5	406	50.4	151	16	Q8GDM4
6	394	48.9	161	16	Q8DHG4
7	363	45.0	161	16	Q8FDM5
8	329.5	40.9	157	16	Q8DZ24
9	272	33.7	152	2	Q8KQPS
10	229	28.4	174	16	Q8GSL2
11	208.5	25.9	171	16	Q8FM14
12	194	24.1	165	16	Q8EUN4
13	180	22.3	89	2	Q8KX06
14	87.5	10.9	571	16	Q8EPR7
15	86	10.7	434	10	Q8LBY9
16	85	10.5	377	16	Q8F721

17	85	10.5	434	10	Q93X74	Q93X74 brassica ra
18	84.5	10.5	429	16	Q8EBJ7	Q8EBJ7 shewanella
19	84.5	10.5	748	16	Q9KZ42	Q9KZ42 streptomyces
20	84	10.4	748	16	Q8X0E8	Q8X0E8 ralsomona s
21	83.5	10.4	410	16	Q8E722	Q8E722 streptococcus
22	83.5	10.4	410	16	Q8E1K5	Q8E1K5 streptococcus
23	83	10.3	373	16	Q8PJEO	Q8PJEO xanthomonas
24	83	10.3	593	5	017849	017849 caenorhabditis
25	82.5	10.2	767	16	Q92T08	Q92T08 rhizobium m
26	82	10.2	939	2	Q9KX10	Q9KX10 corynebacte
27	82	10.2	943	16	Q8NQ98	Q8NQ98 corynebacte
28	81.5	10.1	294	16	Q9KN35	Q9KN35 vibrio chol
29	81.5	10.1	727	16	Q9SEV1	Q9SEV1 rhizobium 1
30	81	10.0	434	10	Q9LPA3	Q9LPA3 arabidopsis
31	80.5	10.0	203	16	Q8FA11	Q8FA11 escherichia
32	80	9.9	209	16	Q8X115	Q8X115 clostridium
33	80	9.9	319	17	Q970V2	Q970V2 sulfobobus
34	80	9.9	319	17	Q970V2	Q970V2 sulfobobus
35	80	9.9	444	10	Q8GVP6	Q8GVP6 oryza sativ
36	79.5	9.9	452	17	Q970M5	Q970M5 sulfobobus
37	79.5	9.9	941	16	Q8FPA8	Q8FPA8 corynebacte
38	79.5	9.9	5206	16	Q8D6P9	Q8D6P9 vibrio vuln
39	79	9.8	903	12	Q64928	Q64928 african hor
40	78.5	9.7	513	16	Q988N0	Q988N0 rhizobium 1
41	78.5	9.7	734	13	Q98C94	Q98C94 rhizobium 1
42	78	9.7	431	13	Q05467	Q05467 gallus gall
43	78	9.7	718	16	Q97EP8	Q97EP8 clostridium
44	78	9.7	1840	13	Q90831	Q90831 gallus gall
45	77.5	9.6	173	16	Q8FPR6	Q8FPR6 corynebacte

ALIGNMENTS

RESULT 1

ID	Q8EBR3	PRELIMINARY:	PRT:	159 AA.
AC	Q8EBR3	01-MAR-2003 (TREMBL:rel. 23, Created)		
DT	01-MAR-2003 (TREMBL:rel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBL:rel. 23, Last annotation update)			
DE	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.			
DN	1SPF OR S03437			
OS	Shewanella oneidensis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;			
OC	Alteromonadaceae; Shewanella.			
OX	NCBI_TaxID=70863;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MR-1;			
RX	MEDLINE=22297686; PubMed=12368813;			
RA	Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,			
RA	Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,			
RA	Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S.,			
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,			
RA	Madupu R., Peterson J.D., Umayam L.A., White C., Wolf A.M.,			
RA	Varatharajan U., Weidman J., Impraim M., Lee K., Berry K., Lee C.,			
RA	Mueller J., Khouri H., Gill J., Uitterlinden T.R., McDonald L.A.,			
RA	Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Frazer C.M.,			
RT	"Genome sequence of the dissimilatory metal ion-reducing bacterium			
RT	Shewanella oneidensis."			
RL	Nat. Biotechnol. 20:1118-1123(2002).			
DR	EMBL; AB015780; AAM56434.1; -			
DR	TIGR; S03437; -			
KW	Complete proteome.			
SQ	SEQUENCE 159 AA; 16995 MW; A16DC82586297501 CRC64;			

Query Match 70.6%; Score 569; DB 16; Length 159;

Best Local Similarity 71.2%; Pred. No. 2,7e-45;

Matches 111; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

2 IRVGMGYDVHRENDGDHILIGSVKIPYEKLEAHSGDDVTLALADAILAALAGDVGK 61

DB 3 IRIGGFVHKEGEPRLILCGEVEYETELVANSDDVVLHAIADALIGAAAGDIGH 62
 QY 62 FPDTPNFKGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIADL 121
 DB 63 FPDTPAAYGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIADL 122
 QY 122 ETDVDFINVKATTEKLGFGKRGKGIYAQAVVLI 157
 DB 123 NADVADINVKATTEKLGFGKRGKGIYAQAVVLI 158

RESULT 2

Q8BC55 PRELIMINARY; PRT; 158 AA.

AC Q8DC59; PRELIMINARY; PRT; 158 AA.
 DB 01-MAR-2003 (TREMELREL. 23, Created)
 DT 01-MAR-2003 (TREMELREL. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.
 GN V11583.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_Taxid=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CNC6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RT "Complete genome sequence of Vibrio vulnificus CNC6."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE016802; AMO1006.1; -
 KM Complete proteome.
 SQ SEQUENCE 158 AA; 17084 MW; 0B888FE050160112 CRC64;

Query Match 68.4%; Score 551; DB 16; Length 158;
 Best Local Similarity 69.0%; Pred. No. 1.3e-43;
 Matches 107; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MTRVGKGVYHRRPNDGHIILGGVYKIPYKGLAHSDGCVLHAIADALIGAAAGDIGH 60
 DB 1 MRIGGFVHKEGEPRLILCGEVEYETELVANSDDVVLHAIADALIGAAAGDIGH 60
 QY 61 FPDTPNFKGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIAD 120
 DB 61 FPDTPAAYGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIAD 120
 QY 121 LETVDVFINVKATTEKLGFGKRGKGIYAQAVVLI 155
 DB 121 LETDININVKATTEKLGFGKRGKGIYAQAVVLI 155

RESULT 3

Q8FB36 PRELIMINARY; PRT; 159 AA.

AC Q8FB36; PRELIMINARY; PRT; 159 AA.
 DB 01-MAR-2003 (TREMELREL. 23, Created)
 DT 01-MAR-2003 (TREMELREL. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (BC
 DE 4.6.1.12).
 GN YGBB OR C3313.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_Taxid=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RA MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Butland V., Plunkett G., III, Redford P., Soesich P.,
 RA Rasko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.;

RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 DR EMBL; AB016765; AAB01762.1; -
 KM Lyase; Complete proteome.
 SQ SEQUENCE 159 AA; 16867 MW; 8FD4573623BA4549 CRC64;

Query Match 66.6%; Score 537; DB 16; Length 159;
 Best Local Similarity 69.5%; Pred. No. 2.6e-42;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVKGVDVHRRPNDGHIILGGVYKIPYKGLAHSDGCVLHAIADALIGAAAGDIGH 61
 DB 1 MRIGGFVHKEGEPRLILCGEVEYETELVANSDDVVLHAIADALIGAAAGDIGH 60
 QY 62 FPDTPNFKGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIADL 121
 DB 61 FPDTPAAYGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIADL 120
 QY 122 ETDVDFINVKATTEKLGFGKRGKGIYAQAVVLI 155
 DB 121 GCHMDVFINVKATTEKLGFGKRGKGIYAQAVVLI 154

RESULT 4

Q8KC25 PRELIMINARY; PRT; 157 AA.

AC Q8KC25; PRELIMINARY; PRT; 157 AA.
 DB 01-OCT-2002 (TREMELREL. 22, Created)
 DT 01-OCT-2002 (TREMELREL. 22, Last sequence update)
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.
 GN I5SP OR CT1601.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobiales; Chlorobiaceae;
 OC Chlorobiaceae; Chlorobium.
 OX NCBI_Taxid=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TJS / ATCC 49652 / DSM 12025;
 RA MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.R., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey R.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radu D.,
 RA Vamathevan J., Khouli H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Teitelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TJS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
 DR EMBL; AB012915; AAM72826.1; -
 DR TIGR; CT1601; -
 DR InterPro; IPR003526; YgbB.
 DR Pfam; PF02542; YgbB; 1.
 DR TIGRPFAMs; TIGR00151; I5SP; 1.
 DR PROSITE; PS01350; I5SP; 1.
 KM Complete proteome.
 SQ SEQUENCE 157 AA; 16923 MW; 4F5A7DCB20C897B CRC64;

Query Match 56.2%; Score 453; DB 16; Length 157;
 Best Local Similarity 54.2%; Pred. No. 1.8e-34;
 Matches 84; Conservative 33; Mismatches 38; Indels 0; Gaps 0;

QY 2 IRVKGVDVHRRPNDGHIILGGVYKIPYKGLAHSDGCVLHAIADALIGAAAGDIGH 61
 DB 1 MRIGGFVHKEGEPRLILCGEVEYETELVANSDDVVLHAIADALIGAAAGDIGH 60
 QY 62 FPDTPNFKGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIADL 121
 DB 61 FPDTPAAYGADSRVLLRHYVGIYKEGKYLVAADVTTIAQAPKPLPVPGKRAANIADL 120

QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 156
 Db 121 GLEINAVSVATTTNEKLGIVGRGEGACAAVCLIE 155

RESULT 5

Q8DGH4 PRELIMINARY; PRT; 451 AA.

AC Q8DGH4; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 DE synthase/4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.
 DE BR1120.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 NCBI_TaxID=29461;
 RX MEDLINE=2224741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Bisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmiller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.B., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL; AE014413; AAN30040.1; -;
 TR TIGR; BR1120; -;
 SQ SEQUENCE 451 AA; 48492 MW; 3BDB7285C056A161 CRC64;

Query Match 50.4%; Score 406; DB 16; Length 451;
 Best Local Similarity 51.3%; Pred. No. 1.7e-29;
 Matches 79; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 2 IRVNGYDVHFRFNDGHIILGVKIPYEKGLFSAHSDGVVLAALDAIIGAAALGDIKH 61
 Db 290 IRTGNGYDVHFRFNDGHIILGVKIPYEKGLFSAHSDGVVLAALDAIIGAAALGDIKH 349
 QY 62 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKLPVPGGRANIAADL 121
 Db 350 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKLPVPGGRANIAADL 409
 QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 155
 Db 410 GLEINAVSVATTTNEKLGIVGRGEGACAAVCLIE 443

RESULT 6

Q8DHC4 PRELIMINARY; PRT; 161 AA.

AC Q8DHC4; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.
 DE TIR2035.
 OS Synecococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 NCBI_TaxID=32046;
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Matsumoto M., Matsumoto M., Nakazaki N.,

RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005376; BAC09587.1; -;
 KW Complete proteome.
 SQ SEQUENCE 161 AA; 17323 MW; CA4C67BB6D02498D CRC64;

Query Match 48.9%; Score 394; DB 16; Length 161;
 Best Local Similarity 45.9%; Pred. No. 5.8e-29;
 Matches 75; Conservative 33; Mismatches 47; Indels 0; Gaps 0;

QY 2 IRVNGYDVHFRFNDGHIILGVKIPYEKGLFSAHSDGVVLAALDAIIGAAALGDIKH 61
 Db 4 IRTGNGYDVHFRFNDGHIILGVKIPYEKGLFSAHSDGVVLAALDAIIGAAALGDIKH 63
 QY 62 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKLPVPGGRANIAADL 121
 Db 64 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKLPVPGGRANIAADL 123
 QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 156
 Db 124 GLEINAVSVATTTNEKLGIVGRGEGACAAVCLIE 158

RESULT 7

Q8FOA5 PRELIMINARY; PRT; 161 AA.

AC Q8FOA5; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE YGDB family protein.
 DE YGDB OR LA3591.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RX MEDLINE=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;
 RA Ren S.;
 RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE011514; AAN50789.1; -;
 KW Complete proteome.
 SQ SEQUENCE 161 AA; 17651 MW; 625A49B3223022C9 CRC64;

Query Match 45.0%; Score 363; DB 16; Length 161;
 Best Local Similarity 45.9%; Pred. No. 4.5e-26;
 Matches 73; Conservative 32; Mismatches 52; Indels 2; Gaps 1;

QY 1 MIRVNGYDVHFRFNDGHIILGVKIPYEKGLFSAHSDGVVLAALDAIIGAAALGDI 58
 Db 1 MYRIGNGYDVHFRFNDGHIILGVKIPYEKGLFSAHSDGVVLAALDAIIGAAALGDI 60
 QY 59 GKPPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKLPVPGGRANIA 118
 Db 61 GKPPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKLPVPGGRANIA 120
 QY 119 ADLETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 157
 Db 121 GLEINAVSVATTTNEKLGIVGRGEGACAAVCLIE 159

RESULT 8

Q8D224 PRELIMINARY; PRT; 157 AA.

AC Q8D224; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE YGDB protein.
 DE YGDB.
 OS Wigglesworthia brevipalpis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wiglesworthia.
 OC NCBI_TaxID=164609;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2297718; PubMed=12219091;
 RA Akman L., Yamashita A., Matanabe H., Oshima K., Shiba T., Hattori M.,
 RA Aksoy S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 RT flies, *Wiglesworthia glossinidia*."
 RL Nat. Genet. 32:402-407 (2002).
 DR EMBL; AB063522; BAC24677.1; -
 KW Complete proteome.
 SQ SEQUENCE 157 AA; 17252 MW; 775DCB1DDAE2884A CRC64;

Query Match 40.9%; Score 329.5; DB 16; Length 157;
 Best Local Similarity 45.2%; Pred. No. 5.8e-23;

Matches 71; Conservative 37; Mismatches 46; Indels 3; Gaps 3;

QY 2 IRVGMGYDVHFRND-GDHIILGCVKIPYKGLBAHSDGVVLAHALADALIGAAALGDIGK 60
 DB 1 MRVGHGFDHFKFKIKYKPLILGCVHPIYCKGVVSHSDGVIIHSILSLGASSLGDIGI 60
 QY 61 HEPDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120
 DB 61 LFPNNDIKYNINSCLLQVWNCIKKK-YKIGNIDVTLFLEYPKISYTNKICICISNC 119
 QY 121 LFTDVPPIVKAATTEKLGFEGRKGLAVQAV-VLIE 156
 DB 120 LKCKTIVINIKSKTMGLGIGKKGIASAVSLLE 156

RESULT 9

Q8RQPS

ID Q8RQPS PRELIMINARY; PRT; 152 AA.

Q8RQPS

DT 01-JUN-2002 (T-EMBLrel. 21, Created)

DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)

DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

DE (EC 4.6.1.12).

GN YGBB.

OS Thermus thermophilus.

OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;

OC Thermus.

OX NCBI_TaxID=274;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=HB8;

RA Kishida H., Wada T., Unzai S., Kuzuyama T., Terada T., Shirouzu M.,

RA Yokoyama S., Tame J.R., Park S.;

RT "2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase from *Thermus*

RT thermophilus HB8."

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB082126; BAB86885.1; -

DR InterPro: IPR003526; YGBB.

DR Pfam: PF02542; YGBB; 1.

DR TIGRPFAM; TIGR00151; ISPF; 1.

DR PROSITE; PS01350; ISPF; 1.

KW Lyase.

SQ SEQUENCE 152 AA; 16520 MW; 01140B44DDC28F3 CRC64;

Query Match 33.7%; Score 272; DB 2; Length 157;

Best Local Similarity 39.4%; Pred. No. 1.3e-17;

Matches 61; Conservative 25; Mismatches 65; Indels 4; Gaps 1;

QY 2 IRVGMGYDVHFRNDGDHIILGCVKIPYKGLBAHSDGVVLAHALADALIGAAALGDIGK 61
 DB 1 MRVGHGFDHFKFKIKYKPLILGCVHPIYCKGVVSHSDGVIIHSILSLGASSLGDIGI 60
 QY 62 HEPDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 121
 DB 61 LFTDVPPIVKAATTEKLGFEGRKGLAVQAV-VLIE 156

QY 122 ETDVDFIVKATTEKLGFEGRKGLAVQAVVLE 156
 DB 121 RLPODRIGLTFKTSGLA-----PSHVQARAVVLD 151

RESULT 10

ID Q8G5L2 PRELIMINARY; PRT; 174 AA.

Q8G5L2

DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, Mecsps.

DE Mecp-synthase.

GN ISPF OR BL0997.

OS Bifidobacterium longum.

OC Bacteria; Actinobacteridae; Actinobacteriales; Bifidobacteriales;

OC Bifidobacteriaceae; Bifidobacterium.

OX NCBI_TaxID=216816;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=NCC 2705;

RX MEDLINE=22294977; PubMed=12381787;

RA Scheil M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,

RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,

RA Pridmore R.D., Arigoni F.;

RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation

RT to the human gastrointestinal tract."

RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).

DR EMBL; AE014722; AAN24805.1; -

KW Complete proteome.

SQ SEQUENCE 174 AA; 17433 MW; F02406B3DBE336B CRC64;

Query Match 28.4%; Score 229; DB 16; Length 174;

Best Local Similarity 37.0%; Pred. No. 1.6e-13;

Matches 64; Conservative 20; Mismatches 67; Indels 22; Gaps 6;

QY 2 IRVGMGYDVHFRNDGDHIILGCVKIPYKGLBAHSDGVVLAHALADALIGAAALGDIGK 46
 DB 6 VLIQGFPAHFRPAGSSRELIMINGLVWPVDSQSDMAAYE-GIEDSDGVVAAHLLI 64
 QY 47 DALIGAAALGDIGKHP-PTDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPXKL 105
 DB 65 DALIAAARLGDIGLFGVGADAHGAGMGIDMLQEVAAHLASNGYTPASASVALIIGKRP 124
 QY 106 MLPVPGKRAIIADLFTDVPF-INVKAATTEKLGFEGRKGLAVQAVVLE 157
 DB 125 -----IGTRRAAABALSAVQCPVSVTNTTTHDKGFTGKSGIAAIAALVYK 173

RESULT 11

ID Q8FM14 PRELIMINARY; PRT; 171 AA.

Q8FM14

DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

GN Conserved hypothetical protein.

GN CE2520.

OS Corynebacterium efficiens.

OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;

OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.

OX NCBI_TaxID=152794;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;

RA Kawaiyayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,

RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,

RA Ueda Y., Sugimoto S.;

RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314."

RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP005222; BAC19330.1; -

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 18076 MW; 524D47207A085E67 CRC64;

Query Match 25.9%; Score 208.5; DB 16; Length 171;
Best Local Similarity 33.1%; Pred. No. 1.2e-11;
Matches 52; Conservative 25; Mismatches 69; Indels 11; Gaps 3;

QY 3 RVGMGYDVHRENDGHIILGVKIPYEKGLAHSDDGVVLAALADALIGAAALDIGKEF 62
DB RVGIASDAHQIEAGKFCMVICLFFDVGDCGEGSDGVVAHAIYDALISASLGDSGFV 78

QY 19 RVGIASDAHQIEAGKFCMVICLFFDVGDCGEGSDGVVAHAIYDALISASLGDSGFV 78
DB 63 PDDDPFKGADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHVGRANITADIE 122

QY 79 GVGREPYDVGSGTQLLEVAELLITRGFTIGVAAQLVQGPFR-----GPRREAOQV 133
DB 123 TDV----DFINVAATTEKLGFEGRKEGIAVQAVLI 155

QY 134 SDILGAPCFHS--ATTDDHMGFTGRGEGRAALATAV 168
DB 134 SDILGAPCFHS--ATTDDHMGFTGRGEGRAALATAV 168

RESULT 12

Q8EUA4 PRELIMINARY; PRT; 165 AA.

AC 08EUA4; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE Putative enzyme of deoxy-xylulose pathway ygbB.

OS Mycoplasma penetrans.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

NCBI_TaxID=28227;

SEQUENCE FROM N.A.

STRAIN=HF-2;

MEDLINE=22354719; PubMed=1246555;

RA Yoshino C, Horino A, Shiba T, Sasaki T, Hattori M, Puroya K,

RA "The complete genomic sequence of Mycoplasma penetrans, an

RT intracellular bacterial pathogen in humans."

Nucleic Acids Res. 30:5293-5300(2002).

DR EMBL; AP004174; BAC44812.1; -

KW Complete proteome.

SQ SEQUENCE 165 AA; 19038 MW; C6851C7BFFBC54 CRC64;

Query Match 24.1%; Score 194; DB 16; Length 165;
Best Local Similarity 31.9%; Pred. No. 2.7e-10;
Matches 51; Conservative 36; Mismatches 57; Indels 16; Gaps 5;

QY 3 RVGMGYDVHRENDGHIILGVKIPYEKGLAHSDDGVVLAALADALIGAAALDIGKEH 61
DB 17 KTGNGIDHDKIKKCKKQRLAGDELDYKITAHSDDGIIHSISALIGALSLDIDGTY 76

QY 62 PDDDPFKGADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHV--PGMRANITAA 119
DB 77 FSDTOSKKNKLDLSLEISCLNKLKQDYSISNVITITCE-----YIIFDIDIDQKS 130

QY 120 DDEFDVD--FINVAKATTEKLGFEGRKEGIAVQAVLIER 157
DB 131 NLEKLENTETISLKAIR-----YEEDKNMIQVTVLILNK 165

RESULT 13

Q8KX06 PRELIMINARY; PRT; 89 AA.

AC 08KX06; 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (Fragment).

OS Synechococcus sp. (strain RCC 7002) (Agmenellum quadruplicatum).

OX NCBI_TaxID=32049;
RN [1]

RP SEQUENCE FROM N.A.

RA Nomura C.T., Persson S., Zhao J., Bryant D.A.;

RT "An Analysis of Forty Genes Encoding Electron Transport Proteins from

RT Synechococcus sp. PCC 7002: A Comparative Study of Electron Transport

RT Proteins from Cyanobacteria and Chloroplasts."

DR Submitted (May-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF381048; AAN03587.1; -

DR InterPro; IPR003526; YGBB.

DR Pfam; PF02542; YGBB; 1.

DR TIGRPFAM; TIGR00151; 1SPF; 1.

FT NON TER

SQ SEQUENCE 89 AA; 9626 MW; 7E5534DC1BF1CE09 CRC64;

Query Match 22.3%; Score 180; DB 2; Length 89;
Best Local Similarity 43.0%; Pred. No. 2.4e-09;
Matches 37; Conservative 18; Mismatches 31; Indels 0; Gaps 0;

QY 70 KGADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHVGRANITADLETVDVFIN 129
DB 1 KGADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHVGRANITADLETVDVFIN 129

QY 130 YKATTEKLGFEGRKEGIAVQAVLI 155
DB 61 YKATTEKLGFEGRKEGIAVQAVLI 155

RESULT 14

Q8EUF7 PRELIMINARY; PRT; 571 AA.

AC 08EUF7; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE Conserved hypothetical protein.

OS Solirubra oncidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

NCBI_TaxID=70863;

SEQUENCE FROM N.A.

STRAIN=MR-1;

MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,

RA Read T.D., Eisen J.A., Seebach R., Ward N., Neche B., Clayton R.A.,

RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,

RA Daboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,

RA Madupu R., Peterson J.D., Umeyam L.A., White O., Wolf A.M.,

RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,

RA Mueller U., Khoult H., Gill J., Utecherback T.R., McDonald L.A.,

RT "Genome sequence of the dissimilatory metal ion-reducing bacterium

RT Shewanella oncidensis."

DR Nat. Biotechnol. 20:1118-1123(2002).

DR EMBL; AE015633; AANS4951.1; -

DR TIGR; S01899; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 571 AA; 64713 MW; 86E9246E7680D81E CRC64;

Query Match 10.9%; Score 87.5; DB 16; Length 571;
Best Local Similarity 24.6%; Pred. No. 12;
Matches 35; Conservative 17; Mismatches 39; Indels 51; Gaps 6;

QY 8 YDVERPNDGHIILGVKIPYEKGLAHSDDGVVLAALADALIGAAALDIGKEH 54
DB 141 YAHILNGKVPNSMLGXTWYVD--EAHLDFDTLKQLESHIRAHVAAINGIAPKTP 197

QY 55 -----LSDIGKHP-----DNDPFGKADSRVLLRHVYGVYKKGKYL 91
DB 198 YHOLKOTLAHYKDLAARIPPTTIPYEVKPGTTSPIKIGIAR-----LKEGGL 248

QY 92 LVNA---DVTIIAQPRLPHV 110
 Db 249 DANATVQDALTVAHAPLIDNT 270

RESULT 15

Q8LHV9 PRELIMINARY; PRT; 434 AA.
 ID Q8LHV9
 AC Q8LHV9;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Monodehydroascorbate reductase (NADH)-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid 1; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_taxid=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RP Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 0:0-0(2002).
 [2]
 RN SEQUENCE FROM N.A.
 RP Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1 - COFACTOR: PAD (BY SIMILARITY).
 DR EMBL; AY086968; AAM64531.1; -;
 DR InterPro; IPR001327; PAD_Pyr_redox.
 DR InterPro; IPR001003; Pyridine_redox_2.
 DR InterPro; IPR001100; Pyr_redox.
 DR Pfam; PF00070; Pyr_redox; 1.
 DR PRINTS; PR00368; FADPR.
 DR PRINTS; PR00411; PYRIDTASE1.
 DR PRINTS; PR00469; PNRDPTASE11.
 DR ProDom; PD000139; PAD_Pyr_redox; 1.
 DR PAD; Flavoprotein; Oxidoreductase.
 KW SEQUENCE 434 AA; 46486 MW; 0DB0C3P6F81D4C0C CRC64;

Query Match 10.7%; Score 86; DB 10; Length 434;

Best Local Similarity 26.7%; Pred. No. 11;

Matches 50; Conservative 22; Mismatches 59; Indels 56; Gaps 8;

QY 22 GGVXIPYE-----KGLAHSGDVVLAADALIGAAALDIGH-----FPPTD 66
 Db 73 GGEKLPESYKQKIEILISTEIVKADLSKSL-VSATGDPFKYQTLITATGTVLRITD 131
 QY 67 PNEKGDSTRVL-----RHVGIYKEKGYLV-----NADVTI 99
 Db 132 FGKKGADSKNIILREIDADKLVBAIKKKGKAVVGGYIGELSAVLINNIDVTM 191
 QY 100 IAGAPKCLHVPGRKANIADLTVDYFNKATL-TEKLGEGR-----KEGTA 148
 Db 192 VFPEPFCMPEL--FTANIAFAETYYTNKGKLIKGTVAAGFTAQPNGEVKEVQINDGRT 249
 QY 149 VQAVVLI 155
 Db 250 LEADIVI 256

Search completed: January 29, 2004, 15:54:10
 Job time : 13.7026 secs

DR N-PSDB; ABK83264.
 XX
 PT New high growth methanotrophic bacterial strain, useful for producing
 PT single cell proteins, grows on a C1 carbon substrate, and comprises a
 PT functional gene encoding in Embden-Meyerhof carbon pathway
 XX
 PS Claim 11; Page 145-147; 157pp; English.
 CC The invention relates to a high growth methanotrophic bacterial strain,
 CC which grows on a C1 carbon substrate e.g. methane and methanol, and
 CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a
 CC gene coding a pyrophosphate dependent phosphotransferase enzyme or a 16S
 CC rRNA. The bacterial strain is useful for the production of single cell
 CC protein and for the biotransformation of a nitrogen-containing compound,
 CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the
 CC production of a feed product comprising a protein, carbohydrates and a
 CC pigment and for reducing oxygen demand, for removing nitrates and
 CC nitrates in methane-containing environments such as landfill, waste
 CC water treatment systems or anywhere that methane, oxygen and nitrates are
 CC present. The bacterial strain of the invention can be used as a
 CC denitrifying agent for the conversion of nitrate or nitrite to nitrous
 CC oxide with methane or methanol as a carbon source. It is also used in the
 CC production of biomass (including proteins, carbohydrates and a wide
 CC variety of pigments (particularly for isoprenoid pigments for the
 CC purpose of generating animal feeds), in production of terpenoid and
 CC carotenoid compounds, useful as pigments and as monomers in polymeric
 CC materials and in production of exopolysaccharides at high levels.
 CC Sequences ABG61551-ABG61590 represent high growth methanotrophic
 CC bacterial strain proteins of the invention.
 XX
 SQ Sequence 544 AA;
 Query Match 100.0%; Score 2788; DB 23; Length 544;
 Best Local Similarity 100.0%; Pred. No. 3e-248;
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTKPFIITGGVSSISGKGIASSLAAILLEBRGLKATITKLDPIYINVDGTMSPFGHGVF 60
 DB 1 MTKPFIITGGVSSISGKGIASSLAAILLEBRGLKATITKLDPIYINVDGTMSPFGHGVF 60
 QY 61 VIEDGAETDLDLGHYERFLKTTMTKNNFTTGOVYEQVLRNKRKGDYLGATVOVPHIND 120
 DB 61 VIEDGAETDLDLGHYERFLKTTMTKNNFTTGOVYEQVLRNKRKGDYLGATVOVPHIND 120
 QY 121 EIKRRVYESAEGQVALIEVGTVGDIESTPLETTRQKGVLEGRRAFLFIHLTVPYIK 180
 DB 121 EIKRRVYESAEGQVALIEVGTVGDIESTPLETTRQKGVLEGRRAFLFIHLTVPYIK 180
 QY 181 SAGELKTPQHSVKELRTIGIOPDLICRSEOPIPASERRKIALFTNVAERKAVISAIIDA 240
 DB 181 SAGELKTPQHSVKELRTIGIOPDLICRSEOPIPASERRKIALFTNVAERKAVISAIIDA 240
 QY 241 DTYRIPLLIREQGLDVLVDQLRNPAAADLSAMERKVDGLTHPTDEVSAIVGRYVDH 300
 DB 241 DTYRIPLLIREQGLDVLVDQLRNPAAADLSAMERKVDGLTHPTDEVSAIVGRYVDH 300
 QY 301 TDYVKSINELTHAGITHRHXVOISYIDSTTSABGTAKIKNDALIVGGREGRVGRK 360
 DB 301 TDYVKSINELTHAGITHRHXVOISYIDSTTSABGTAKIKNDALIVGGREGRVGRK 360
 QY 361 ISTVFARENKIPYLGICLQMOQSAVIEFARNVVGLEGAHSTEFLLPKSPFVIGLITEND 420
 DB 361 ISTVFARENKIPYLGICLQMOQSAVIEFARNVVGLEGAHSTEFLLPKSPFVIGLITEND 420
 QY 421 EAGEIVTRDSDSLGCMRGLGAKCKLKAADSLAFOLYQKOVYTERRRHREYFNQYLKOL 480
 DB 421 EAGEIVTRDSDSLGCMRGLGAKCKLKAADSLAFOLYQKOVYTERRRHREYFNQYLKOL 480
 QY 481 EAAGKFSKSGSLDGRVLEIIELEHWPFLACQHPERTSTPRNGHLLFSGFVAAAKHNT 540
 DB 481 EAAGKFSKSGSLDGRVLEIIELEHWPFLACQHPERTSTPRNGHLLFSGFVAAAKHNT 540
 QY 541 QGTA 544

DB 541 QGTA 544
 ||||
 RESULT 2
 ID AAE22306
 XX AAE22306 standard; Protein; 544 AA.
 AC AAE22306;
 XX
 DT 25-JUL-2002 (First entry)
 XX
 DS Methylomonas 16a sp. CTP synthase (PyrG) enzyme.
 XX
 KW Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;
 KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;
 KW aquaculture; enzyme; CTP synthase; Pyrg.
 XX
 OS Methylomonas 16a sp.
 XX
 MO200218617-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 04-SEP-2001; 2001WO-US27420.
 XX
 PR 01-SEP-2000; 2000US-223858P.
 PR 01-SEP-2000; 2000US-223907P.
 XX
 PA (DUPO) DU PONT DE MEMOURS & CO E. I.
 XX
 PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;
 PI Odum JM, Picataggio SK, Rouviere PE;
 XX
 MP1: 2002-351711/38.
 DR N-PSDB; AAD35504.
 XX
 PT Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by
 PT using microorganisms having a nucleic acid molecule encoding enzymes in
 PT the carotenoid biosynthetic pathway and which metabolize single carbon
 PT substrates
 XX
 PS Claim 46; Page 120-122; 156pp; English.
 CC The invention relates to a method for producing carotenoid compounds.
 CC The method comprises a transformed metabolizing host cell, comprising
 CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
 CC encoding an enzyme in the carotenoid biosynthetic pathway, under the
 CC control of regulatory sequences, and contacting the host cell with carbon
 CC substrate to produce a carotenoid compound. The method is useful for
 CC producing carotenoid compounds such as anthraxanthin and astaxanthin, by
 CC using microorganisms having a nucleic acid molecule encoding enzymes in
 CC the carotenoid biosynthetic pathway and which metabolize single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Methylomonas 16a sp. CTP synthase (PyrG) enzyme used in the invention.
 XX
 SQ Sequence 544 AA;
 Query Match 100.0%; Score 2788; DB 23; Length 544;
 Best Local Similarity 100.0%; Pred. No. 3e-248;
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTKPFIITGGVSSISGKGIASSLAAILLEBRGLKATITKLDPIYINVDGTMSPFGHGVF 60
 DB 1 MTKPFIITGGVSSISGKGIASSLAAILLEBRGLKATITKLDPIYINVDGTMSPFGHGVF 60
 QY 61 VIEDGAETDLDLGHYERFLKTTMTKNNFTTGOVYEQVLRNKRKGDYLGATVOVPHIND 120
 DB 61 VIEDGAETDLDLGHYERFLKTTMTKNNFTTGOVYEQVLRNKRKGDYLGATVOVPHIND 120

QY 121 EIKRRVYESAEKGDVALIEVGTVGDIESTPLFETIRQMGVGLGRDRLFIHLTLVPYIK 180
 Db 121 EIKRRVYESAEKGDVALIEVGTVGDIESTPLFETIRQMGVGLGRDRLFIHLTLVPYIK 180
 QY 181 SAGELKTKPTQHSYKELRTIGIQDIIICRSEQIPASERKIALFTNVAEKAVISIDA 240
 Db 181 SAGELKTKPTQHSYKELRTIGIQDIIICRSEQIPASERKIALFTNVAEKAVISIDA 240
 QY 241 DTIYRIPLLRBOGLDVLVVDQLRLDVPADLSAMEKVDGLTHPDEVSAIIVGKYVDH 300
 Db 241 DTIYRIPLLRBOGLDVLVVDQLRLDVPADLSAMEKVDGLTHPDEVSAIIVGKYVDH 300
 QY 301 TDVAKSLNEALIHAGIHTHRKVOISYIDSETTEAGTAKLKNVDAIIVPGFGEVGEK 360
 Db 301 TDVAKSLNEALIHAGIHTHRKVOISYIDSETTEAGTAKLKNVDAIIVPGFGEVGEK 360
 QY 361 ISTVRPAERKIPYLGLICMGQSAVIEFARNVGLBGHSTFELPSPHVIGLITEMND 420
 Db 361 ISTVRPAERKIPYLGLICMGQSAVIEFARNVGLBGHSTFELPSPHVIGLITEMND 420
 QY 421 EAGELVTRDEDSDLGTMRLGAQKRLKADSLAFOLYOKDVIITERHRRHYEFNNQYLKOL 480
 Db 421 EAGELVTRDEDSDLGTMRLGAQKRLKADSLAFOLYOKDVIITERHRRHYEFNNQYLKOL 480
 QY 481 EAAQMKFSGKSLDGLVEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540
 Db 481 EAAQMKFSGKSLDGLVEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540
 QY 541 QGTA 544
 Db 541 QGTA 544

RESULT 3
 ID AAB0330 standard; Protein: 544 AA.
 AC AAB0330;
 DT 15-JUL-2002 (first entry)
 DE Methylomonas 16a ORF6 pyrg (CTP synthase enzyme) protein sequence.
 XX
 KW Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive; kerateneid; pigment; flavour; fragrance; open reading frame 6; ORF6;
 KM pyrg; CTP synthase enzyme.
 XX
 OS Methylomonas sp.
 XX
 PN WC0200220733-A2.
 PD 14-MAR-2002.
 XX
 PF 29-AUG-2001; 2001MO-US26852.
 XX
 PR 01-SEP-2000; 2000US-229907P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cheng Q, Kofas M, Norton KC, Odom JM, Picataggio SK, Rouviere PE;
 PI Schenzle A, Tomb J;
 XX
 DR WPI; 2002-383051/41.
 DR N-PSDB; ABR50086.
 XX
 PT Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
 PT isolated from Methylomonas 16a, useful for the production of isoprenoid
 PT compounds -
 XX
 PS Claim 4; Page 74-76; 84pp; English.
 CC The present invention relates to a new nucleic acid molecule encoding
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.

CC The invention is useful for obtaining a nucleic acid molecule
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the
 CC microbial production of isoprenoid compounds. The molecules of the
 CC invention are also useful for regulating isoprenoid biosynthesis in an
 CC organism and for producing recombinant organisms for producing various
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,
 CC for the production of kerateneids and their derivatives, isoprenoid
 CC intermediates, and as pure products useful as pigments, flavours and
 CC fragrances. The present amino acid sequence represents the Methylomonas
 CC 16a open reading frame 6 (ORF6) pyrg (CTP synthase enzyme) protein of
 CC the invention, as described above.

Sequence - 544 AA:
 Query Match 100.0%; Score 2788; DB 23; Length 544;
 Best local Similarity 100.0%; Pred. No. 3e-248;
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKFIFITGGVSSLGKGIASSLAILEDRLKVTITLADPYINVPQMGSPFGHGVF 60
 Db 1 MTKFIFITGGVSSLGKGIASSLAILEDRLKVTITLADPYINVPQMGSPFGHGVF 60
 QY 61 VTEKCASTYLDLGHYERFLKTTTKKNNFTTQVYEOVLRNKRKDYLGATVQVPIHID 120
 Db 61 VTEKCASTYLDLGHYERFLKTTTKKNNFTTQVYEOVLRNKRKDYLGATVQVPIHID 120
 QY 121 EIKRRVYESAEKGDVALIEVGTVGDIESTPLFETIRQMGVGLGRDRLFIHLTLVPYIK 180
 Db 121 EIKRRVYESAEKGDVALIEVGTVGDIESTPLFETIRQMGVGLGRDRLFIHLTLVPYIK 180
 QY 181 SAGELKTKPTQHSYKELRTIGIQDIIICRSEQIPASERKIALFTNVAEKAVISIDA 240
 Db 181 SAGELKTKPTQHSYKELRTIGIQDIIICRSEQIPASERKIALFTNVAEKAVISIDA 240
 QY 241 DTIYRIPLLRBOGLDVLVVDQLRLDVPADLSAMEKVDGLTHPDEVSAIIVGKYVDH 300
 Db 241 DTIYRIPLLRBOGLDVLVVDQLRLDVPADLSAMEKVDGLTHPDEVSAIIVGKYVDH 300
 QY 301 TDVAKSLNEALIHAGIHTHRKVOISYIDSETTEAGTAKLKNVDAIIVPGFGEVGEK 360
 Db 301 TDVAKSLNEALIHAGIHTHRKVOISYIDSETTEAGTAKLKNVDAIIVPGFGEVGEK 360
 QY 361 ISTVRPAERKIPYLGLICMGQSAVIEFARNVGLBGHSTFELPSPHVIGLITEMND 420
 Db 361 ISTVRPAERKIPYLGLICMGQSAVIEFARNVGLBGHSTFELPSPHVIGLITEMND 420
 QY 421 EAGELVTRDEDSDLGTMRLGAQKRLKADSLAFOLYOKDVIITERHRRHYEFNNQYLKOL 480
 Db 421 EAGELVTRDEDSDLGTMRLGAQKRLKADSLAFOLYOKDVIITERHRRHYEFNNQYLKOL 480
 QY 481 EAAQMKFSGKSLDGLVEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540
 Db 481 EAAQMKFSGKSLDGLVEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540
 QY 541 QGTA 544
 Db 541 QGTA 544

RESULT 4
 ID AAB6340 standard; Protein: 542 AA.
 AC AAB6340;
 DT 14-FEB-2002 (first entry)
 DE Pseudomonas aeruginosa cellular proliferation protein #330.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Pseudomonas aeruginosa.

```

XX  NO2001:70955-A2.
XX  27-SEP-2001.
XX  21-MAR-2001; 2001WO-US09180.
XX  21-MAR-2000; 2000US-191078P.
XX  23-MAY-2000; 2000US-206848P.
XX  26-MAY-2000; 2000US-207727P.
XX  27-OCT-2000; 2000US-242578P.
XX  27-NOV-2000; 2000US-253625P.
XX  22-DEC-2000; 2000US-257931P.
XX  16-FEB-2001; 2001US-269308P.
XX  (ELIT-) ELITRA PHARM INC.
XX  Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX  Yamamoto RT, Xu HH;
XX  WPI; 2001-611495/70.
XX  N-PSDB; AAS54199.
XX  New polynucleotides for the identification and development of
XX  antibiotics, comprise sequences of antisense nucleic acids -
XX  Example 3; Seq ID No 11933; 511pp; English.
XX  The invention relates to antisense inhibitors of genes essential to
XX  prokaryotic cellular proliferation, their use in identifying the
XX  genes, their use in the discovery of novel antibiotics, the essential
XX  genes themselves and the encoded proteins. The prokaryotes used are
XX  Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX  pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX  invention is also useful for the identification of potential new targets
XX  for antibiotic development. The antisense nucleic acids can also be used
XX  to identify proteins used in proliferation, to express these proteins,
XX  and to obtain antibodies capable of binding to the expressed proteins.
XX  The proteins can be used to screen compounds in rational drug discovery
XX  programmes. The antisense nucleic acid sequence is also useful to screen
XX  for homologous nucleic acids which are required for cell proliferation in
XX  a wide variety of organisms. The present sequence represents an
XX  essential prokaryotic cellular proliferation protein.
XX  Note: The sequence data for this patent did not form part
XX  of the printed specification, but was obtained in electronic
XX  format directly from WIPO at
XX  ftp.wipo.int/pub/published_pct_sequences.
XX  Sequence 542 AA.
XX  Query Match 72.0%; Score 2008; DB 22; Length 542;
XX  Best Local Similarity 71.2%; Pred. No. 3e-176;
XX  Matches 383; Conservative 64; Mismatches 91; Indels 0; Gaps 0;
XX  1 MTKEFTTGGVSSLSKGIASSLAALTEBRGKVTITLDPYINVDPGTMSFEGHGVF 60
XX  1 MTKYIVTGGVSSLSKGIASSLAALTEBRGKVTITLDPYINVDPGTMSFEGHGVF 60
XX  61 VTODGAETDLDLGHYERFLKTTWKXNNFTTGGVYEQVLNENKGDYLGATYQVPIPHITD 120
XX  61 VTEDGAETDLDLGHYERFLKTTWKXNNFTTGGVYEQVLNENKGDYLGATYQVPIPHITD 120
XX  61 VTODGAETDLDLGHYERFLKTTWKXNNFTTGGVYEQVLNENKGDYLGATYQVPIPHITD 120
XX  121 EIKRRRYVSEAGKDVLLIIVGTVGDIIEIPELETIRMGVDELGRRAFLHLLTVPIYK 180
XX  121 EIKRRRIKGGADVALVEIGTVGDIIEIPELETIRMGVDELGRRAFLHLLTVPIYK 180
XX  121 EIKRRRIKGGADVALVEIGTVGDIIEIPELETIRMGVDELGRRAFLHLLTVPIYK 180
XX  181 SAGELTKXPTQHSVKELRTIGIOPDILICRSEQPIPASRRKIALFTYNAEKVISAIDA 240
XX  181 TAGETKXPTQHSVKELRTIGIOPDILICRSEQPIPASRRKIALFTYNAEKVISAIDA 240
XX  241 DTIRYRPLLRBGLDVLVVDQRLDVPADLSAMKRYVDGLTHPIDEXISAIYKRYVDH 300
XX  241 DTIRYRIPSVLHAQGLDVLVVDQRLDVPADLSAMKRYVDGLTHPIDEXISAIYKRYVDH 300

```

```

XX  301 TDAYSLINELIAGIHTRHRVQISYIDSTIEAGTAKLNKVDAILVPGSGENGVECK 360
XX  301 IDAYKSLIEAMTHAGIOSRTKYNLFIDSEIDEGQSTILBEGVDAILVPGSGENGVECK 360
XX  361 ISTVAPARENKIPYIGICLQMSAVIEFARNVVGEGASHTREFFKSPHPVGLITEMMD 420
XX  361 ISTVAPARENKIPYIGICLQMSAVIEFARNVVGEGASHTREFFKSPHPVGLITEMMD 420
XX  421 EAGELVTRDESDLDGTVNRIGAKCRLRADSLAFQLYOKDVIATERHRRRYEFBNQYLKOL 480
XX  421 AGGATEIRTEASDLGTVNRIGAKCRLRADSLAFQLYOKDVIATERHRRRYEFBNQYLKOL 480
XX  481 EAGMKTFGSKSLDGLVLIIELEPHRPWLACQFHEPFTSTPRNKALRSGFVEAAKH 538
XX  481 EQAGIKISGSGDGLVEVEAPBHPWVACQFHEPFTSTPRDGHPLFSGFVNAALKY 538
XX  Db
XX  RESULT 5
XX  AAU34682
XX  ID AAU34682 standard; Protein; 545 AA.
XX  AC AAU34682;
XX  DT 14-FEB-2002 (first entry)
XX  DE E. coli cellular proliferation protein #263.
XX  KM Antisense; prokaryotic cellular proliferation protein;
XX  KM antibiotic; antibacterial; drug design.
XX  OS Escherichia coli.
XX  PN WO200170955-A2.
XX  PD 27-SEP-2001.
XX  PE 21-MAR-2001; 2001WO-US09180.
XX  PR 21-MAR-2000; 2000US-191078P.
XX  PR 23-MAY-2000; 2000US-206848P.
XX  PR 26-MAY-2000; 2000US-207727P.
XX  PR 23-OCT-2000; 2000US-242578P.
XX  PR 27-NOV-2000; 2000US-253625P.
XX  PR 22-DEC-2000; 2000US-257931P.
XX  PR 16-FEB-2001; 2001US-269308P.
XX  PA (ELIT-) ELITRA PHARM INC.
XX  PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX  PI Yamamoto RT, Xu HH;
XX  WPI; 2001-611495/70.
XX  N-PSDB; AAS52541.
XX  New polynucleotides for the identification and development of
XX  antibiotics, comprise sequences of antisense nucleic acids -
XX  Example 3; Seq ID No 10275; 511pp; English.
XX  The invention relates to antisense inhibitors of genes essential to
XX  prokaryotic cellular proliferation, their use in identifying the
XX  genes, their use in the discovery of novel antibiotics, the essential
XX  genes themselves and the encoded proteins. The prokaryotes used are
XX  Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX  pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX  invention is also useful for the identification of potential new targets
XX  for antibiotic development. The antisense nucleic acid sequence is also
XX  used to identify proteins used in proliferation, to express these proteins,
XX  and to obtain antibodies capable of binding to the expressed proteins.
XX  The proteins can be used to screen compounds in rational drug discovery
XX  programmes. The antisense nucleic acid sequence is also useful to screen
XX  for homologous nucleic acids which are required for cell proliferation in

```

CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 545 AA;

Query Match 68.7%; Score 1916; DB 22; Length 545;
Best Local Similarity 67.7%; Pred. No. 9.3e-168;

Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

DB 3 TNYITFGGVSSLSGKGIASSLAALIEDRGKVTITKLDPIYNDPQTMSPFOHGEVAV 61
QY 2 TKRIFITGGVSSLSGKGIASSLAALIEDRGKVTITKLDPIYNDPQTMSPFOHGEVAV 61
DB 3 TNYITFGGVSSLSGKGIASSLAALIEDRGKVTITKLDPIYNDPQTMSPFOHGEVAV 62
QY 62 TEDGAEITDLDLGHYERFLTKMTKNNFTTQGVYEQVLRNKRKDYLGATVOYIPIHIDE 121
DB 63 TEDGAEITDLDLGHYERFLTKMTKNNFTTQGVYEQVLRNKRKDYLGATVOYIPIHIDE 122
QY 122 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLALFIHLTLVPYIKS 181
DB 123 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLALFIHLTLVPYIKS 182
QY 182 AGSLKTPTOHSHVKBRTTIGIQPDILICRSEOPIPASERRKIALFTNAEKAVISAIADAD 241
DB 183 SGSEVTKPTQSHVKBRTTIGIQPDILICRSDRAVPANERAKIALFCNVEPKAVISLKDVD 242
QY 242 TITRIFLLRGGDLVDVLDLADVPADLSAMENYVQGLTHPTDEVSIATVYKRVYHT 301
DB 243 STIKIRPLKSGDLVDYICRPSFLNCEPNAISMEVIEEAPNSEVITIGWGXKIEAP 302
QY 302 DAYKSLMALIHAGIHTHRAVOISYIDSETIEAGTAKLKNVDAILVPGFGERGEGKI 361
DB 303 DAYKSVIALKAGHGLKNRVSVINIKLDSQVETRGVILKGLDAILVPGFGERGEGKI 362
QY 362 STVRPARENKIPLYGICLGMOSAVIEPARNVGLBGAHSTETPLPSPHAVIGLITEMMD 421
DB 363 TTRAPARENKIPLYGICLGMOSAVIEPARNVGLBGAHSTETPLPSPHAVIGLITEMMD 422
QY 422 AGELVTRDESDSLGGMRLGAOKCRLKADSLAQLOKQVITERHHRERFENNQVILKOLE 481
DB 423 NGWVETSESDSLGGMRLGAOKCRLKADSLAQLOKQVITERHHRERFENNQVILKOLE 482
QY 482 AAGKETSGLKSLDRGLVEIIELEPHWFLACQEPHETSTPRNGHALFSGFVEAAK 537
DB 483 DAGLRVAGSGDDQVLVEIIELEPHWFLACQEPHETSTPRNGHALFSGFVEAAK 538

RESULT 6

AAU55582 standard; Protein; 545 AA.

XX AC AAU35582;

XX 14-FEB-2002 (first entry)

XX Haemophilus influenzae cellular proliferation protein #223.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Haemophilus influenzae.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001MO-US09180.

XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253252P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto R, Xu HH;

DR WPI; 2001-611495/70.
DR N-PSDB; AAS53441.

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11175; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 545 AA;

Query Match 68.5%; Score 1908.5; DB 22; Length 545;
Best Local Similarity 67.5%; Pred. No. 4.6e-167;

Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

QY 2 TKRIFITGGVSSLSGKGIASSLAALIEDRGKVTITKLDPIYNDPQTMSPFOHGEVAV 61
DB 3 TNYITFGGVSSLSGKGIASSLAALIEDRGKVTITKLDPIYNDPQTMSPFOHGEVAV 62
QY 62 TEDGAEITDLDLGHYERFLTKMTKNNFTTQGVYEQVLRNKRKDYLGATVOYIPIHIDE 121
DB 63 TEDGAEITDLDLGHYERFLTKMTKNNFTTQGVYEQVLRNKRKDYLGATVOYIPIHIDE 122
QY 122 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLALFIHLTLVPYIKS 181
DB 123 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLALFIHLTLVPYIKS 182
QY 182 AGSLKTPTOHSHVKBRTTIGIQPDILICRSEOPIPASERRKIALFTNAEKAVISAIADAD 241
DB 183 AGSLKTPTOHSHVKBRTTIGIQPDILICRSEOPIPASERRKIALFTNAEKAVISAIADAD 242
QY 242 TITRIFLLRGGDLVDVLDLADVPADLSAMENYVQGLTHPTDEVSIATVYKRVYHT 301
DB 243 STIKIRPLKSGDLVDYICRPSFLNCEPNAISMEVIEEAPNSEVITIGWGXKIEAP 302
QY 302 DAYKSLMALIHAGIHTHRAVOISYIDSETIEAGTAKLKNVDAILVPGFGERGEGKI 361
DB 303 DAYKSVIALKAGHGLKNRVSVINIKLDSQVETRGVILKGLDAILVPGFGERGEGKI 362
QY 362 STVRPARENKIPLYGICLGMOSAVIEPARNVGLBGAHSTETPLPSPHAVIGLITEMMD 421
DB 363 TTRAPARENKIPLYGICLGMOSAVIEPARNVGLBGAHSTETPLPSPHAVIGLITEMMD 422

QY 422 AGELVTRDESDLDGGTMRGAQKCRKLKADSLAFOLYOKDVITERHRRRYEFNNQYLKOLE 481
 DB 423 EGNVTRTDSDDGGTMRGAQCHLVSGSRARELVKERTIEERHRRRYEVNNTLLPOLIE 482
 QY 482 AAGMKRSGKSLDGLVETIELPEHPWFLACQFHEPFTSPRNGHALPSCGVAA-AKH 539
 DB 483 KAGELKVTGLSADKKVLEIIEVPHMPFVACQFHEPFTSPRNGHPLFAGVKAVERNK 541
 RESULT 7
 ABP76969
 ID ABP76969 standard; Protein; 544 AA.
 AC ABP76969;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 468.
 DE N. gonorrhoeae; infection; vaccine; gene therapy.
 KW Antibacterial;
 OS Neisseria gonorrhoeae.
 PN WO200279243-A2.
 PD 10-OCT-2002.
 PF 12-FEB-2002; 2002WO-IB02069.
 PR 12-FEB-2001; 2001GB-0003424.
 PA (CHIR-) CHIRON SPA.
 PI Fontana MR, Pizsa M, Masignani V, Monaci E;
 DR N-PSDB; AB237939.
 DR N-PSDB; AB237939.
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection
 PS Disclosure; Page 222; 815pp; English.
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 SQ Sequence 544 AA;
 Query Match 67.8%; Score 1889.5; DB 24; Length 544;
 Best Local Similarity 66.2%; Pred. No. 2.6e-165;
 Matches 359; Conservative 78; Mismatches 102; Indels 3; Gaps 2;

QY 241 DTVIRIPILLRBOGDDLVQDLRDVPAADLSANEXVVDGLTPTDEVSLAVGVYVDH 300
 DB 241 DSTYECPEMLHDQGDIDNITTEQLQNLVQCADLITANKKIVHAYKAKHIVKILAMGVKYD 300
 QY 301 TDAYSINLEALIHAGIHTFRHKVQISYIDSETEAE--GTAKLKNVDAILVPGGEGEVE 358
 DB 301 TESYKSLIEALKHAGIHETEDVQITFVDSESEKKNKGDVSLKDDAILVPGGEGEVE 360
 QY 359 GKISTVFARRENKIYLLICGMSAVIEFARNVYGLGASHTERPLSPHVLITLTEM 418
 DB 361 GKIAAVFARRENKVVYLLICGMSQVTLFVARDVAGLKGASTERTDLKCAAPVALLTEM 420
 QY 419 MDEAGELVTRDESDLDGGTMRGAQKCRKLKADSLAFOLYOKDVITERHRRRYEFNNQYLK 478
 DB 421 QTAQGSVETROESADLDGGTMRGAQVELELKGSLAVKLYGSGHIERHRRRYEVNNTVYS 480
 QY 479 QLEAAGKTFSGKSLD-GLVETIELPEHPWFLACQFHEPFTSPRNGHALPSCGVAAAK 537
 DB 481 ALEQGLVYIGVSGAGRELVETIELPHMPFVACQFHEPFTSPRNGHPLFAGVKAALN 540
 QY 538 HK 539
 DB 541 NK 542
 RESULT 8
 AAD38272
 ID AAD38272 standard; Protein; 511 AA.
 AC AAD38272;
 DT 14-FEB-2002 (first entry)
 DE Salmonella typhi cellular proliferation protein #163.
 DE Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Salmonella typhi.
 PN WO200170955-A2.
 PD 27-SEP-2001.
 PF 21-MAR-2001; 2001WO-US09180.
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 PA (ELIT-) ELITRA PHARM INC.
 PI Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Traxick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS56131.
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PS Example 3; Seq ID No 13865; 511pp; English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

CC pneumoniae, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programs. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_gct_sequences.

XX Sequence 511 AA;

Query Match 64.0%; Score 1783; DB 22; Length 511;
 Best Local Similarity 67.4%; Pred. No. 1.6e-155;
 Matches 341; Conservative 64; Mismatches 101; Indels 0; Gaps 0;

```

QY 2 TKRFIFITGVVSSLGKGIASSLAILEDRLKVTITKLPYINVDPTGMSPPHGEVYF 61
DB 3 TNNIFVITGVVSSLGKGIASSLAILEDRLKVTITKLPYINVDPTGMSPPHGEVYF 62
QY 62 TEDGAEITDLDGHERLFTKMTKNNFTTGOVYEOVLNBRKGDYIGATVOYIPIHTD 121
DB 63 TEDGAEITDLDGHERLFTKMTKNNFTTGOVYEOVLNBRKGDYIGATVOYIPIHTD 122
QY 122 IKRRVYASAGKQVALIEVGTVGDIESLPFLFTIRMGVELGRDALFIHLTVPIYS 181
DB 123 IKRRVYASAGKQVALIEVGTVGDIESLPFLFTIRMGVELGRDALFIHLTVPIYS 182
QY 182 AGELTKPTQHSVKELETTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISAI 241
DB 183 AGELTKPTQHSVKELETTIGIOPDILICRSDRAVPANERKIALFCNVEKAVISAI 242
QY 242 TIYRPLILREOGDLVDQLRLDVPADLSAMEKVVNGLTHPTDEVSIAIVGKYVDHT 301
DB 243 SIYKIFGLIKSGDLVDYICRFSINCEANLSEMQIYIEANPAGEVITGMVGKYLEHP 302
QY 302 DAKSYNEALIHAGIHTRHKVOISYIDSETIEABGTAKLKNVDAILVPGFGEVGEKI 361
DB 303 DAKSYNEALIHAGIHTRHKVOISYIDSETIEABGTAKLKNVDAILVPGFGEVGEKI 362
QY 362 STVRFARENKIPLYGICLGMOSAVIEFARNVGLGASHTEFLPKSPHPVIGLITEMDB 421
DB 363 ATARVARENKIPLYGICLGMQVALIEFARNVAGMDANSTEFVPGCKEYVALITERRB 422
QY 422 AGELVTRDESDSLGTMRLGAOKRLKADSLAFOLYOKDYITERHRRYEPNNQYIKOLE 481
DB 423 DGNVAVRSEKSDLGTMRLGAOCOSDLSLVOLYGASTIVERRHRRYEVNNMLKOLE 482
QY 482 AAGMKFSKSLDGLRLVEIIELEPHPW 507
DB 483 AAGKXKXGSGDDQLVEIIEVPHMX 508

```

RESULT 9
 AAR15642
 ID AAR15642 standard; Protein; 535 AA.

XX AAR15642;
 XX
 AC AAR15642;
 XX
 DT 25-MAR-2003 (updated)
 DT 05-FEB-1992 (first entry)
 XX
 DE CTP synthetase.
 XX
 KM Cytidine.
 XX
 OS *Bacillus natto* C-1.
 XX

PN JP03228689-A.
 XX
 XX 09-OCT-1991.
 PD
 XX
 XX 05-FEB-1990; 90JP-0025574.
 PF
 XX
 XX 05-FEB-1990; 90JP-0025574.
 PR
 XX
 XX (ASAH) ASAH CHEM IND CO LTD.
 PA
 XX
 XX WPI; 1991-343171/47.
 DR
 XX N-PSDB; AAQ14668.
 DR

PT Prod. of cytidine by fermentation with increased yield - comprises
 PT transforming a suitable microorganism with a recombinant DNA contg.
 PT CTP synthetase gene and incubating with a uracil cpd.
 PT

PS Disclosure; Fig 2; 8pp; Japanese.

CC Microorganisms transformed with the gene can be used to produce
 CC cytidine for use as a starting cpd. for prodn. of drugs.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC

XX Sequence 535 AA;

Query Match 58.2%; Score 1622.5; DB 12; Length 535;
 Best Local Similarity 57.0%; Pred. No. 1.1e-140;
 Matches 310; Conservative 94; Mismatches 127; Indels 13; Gaps 3;

```

QY 1 MKRFIFITGVVSSLGKGIASSLAILEDRLKVTITKLPYINVDPTGMSPPHGEVYF 60
DB 1 MKRYIFVITGVVSSLGKGIASSLAILEDRLKVTITKLPYINVDPTGMSPPHGEVYF 60
QY 61 VEDGAEITDLDGHERLFTKMTKNNFTTGOVYEOVLNBRKGDYIGATVOYIPIHTD 120
DB 61 VEDGAEITDLDGHERLFTKMTKNNFTTGOVYEOVLNBRKGDYIGATVOYIPIHTD 120
QY 121 IKRRVYASAG--GDOVALIEVGTVGDIESLPFLFTIRMGVELGRDALFIHLTVPIY 178
DB 121 IKRRVYASAGKETAADVITTEIGTVGDIESLPFLFTIRMGVELGRDALFIHLTVPIY 180
QY 179 ITSAGELTKPTQHSVKELETTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISAI 238
DB 181 IKAAGELTKPTQHSVKELETTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISAI 240
QY 239 DADTITRIPLILREOGDLVDQLRLDVPADLSAMEKVVNGLTHPTDEVSIAIVGKYV 298
DB 241 DADNITSIFLPLKOKGDLKLVCEHMKLACKRAENGEKELVAKVSNLSQITITIGVGYV 300
QY 299 DHTDAVKSLNEALIHAGIHTRHKVOISYIDSETIEABGTAKLKNVDAILVPGFGEV 357
DB 301 ELDPDAVYISVESELRAHAGYAFPTDVYVKKWINAIEVTENNIAELTSGTDEGIIVGSGDRGV 360
QY 368 BCKITVAFARENKIPLYGICLGMOSAVIEFARNVGLGASHTEFLPKSPHPVIGLITE 417
DB 361 BCKITVAFARENKIPLYGICLGMQVALIEFARNVAGMDANSTEFVPGCKEYVALITERR 420
QY 418 WMDENAGBLVTRDESDSLGTMRLGAOKRLKADSLAFOLYOKDYITERHRRYEPNNQYL 477
DB 421 QMDV-----EDLGGTIRLGLYPCKEBGTVAPEVYODEVYVEKRRHRYEPNNFR 470
QY 478 KOLEAAGKKGSGKSLDGLRLVEIIELEPHPWFLACQFHEEFTSTPANGALTSQVPEAAK 537
DB 471 QOMESQGFVFGTSPDGLVEIIELEKHPWVAVQAFHEFYSRPTROPPLRGFGIAGASV 530
QY 538 HKTQ 541
DB 531 AANQ 534

```

RESULT 10
 AAB48439
 ID AAB48439 standard; Protein; 532 AA.

XX AB048439;
 AC 05-FEB-2002 (first entry)
 DT Listeria monocytogenes protein #1143.
 DE Anticardial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.
 XX Listeria monocytogenes.
 OS
 XX MO200177335-A2.
 XX 18-OCT-2001.
 XX 11-APR-2001; 2001MO-FR01118.
 XX 11-APR-2000; 2000FR-0004629.
 XX (INSP) INST PASTEUR.
 XX Buchrieser C, Frangoul L, Couve E, Rusniok C, Feihl H, Dehoux P,
 PI Dussurget O, Chetoui M, Nedjati H, Glaser P, Kunst F, Cossart P,
 PI Dominguez-Bernal G, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Bagueri F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Medunovic E, De Pablos B, Weiland J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;
 XX MPI; 2002-010914/01.
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX Claim 6; SEQ ID No 1144; 192bp; French.
 XX The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see AB030641). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 532 AA;
 SQ
 Query Match 57.4%; Score 1600.5; DB 21; Length 532;
 Best Local Similarity 55.2%; Pred. No. 1.2e-138;
 Matches 299; Conservative 104; Mismatches 126; Indels 13; Gaps 3;

DB 121 ELKORVFAAAMTNSDITITTEIGTGVDSLPPELHAIKQIKDVGAEENVYIHITLIPY 180
 QY 179 IKSAGELTKPTQSHSKELRTTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAI 238
 DB 181 IKAAGEKTKPTQSHSKELRTTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAI 240
 QY 239 DADTYRIPILREGLDLDVVDOLDPVPAADLSAMEKVVDGLPHPTDEVSIAIVGVXV 298
 DB 241 DEETLYNVPVLSIQKMDIVLEHILQEAPOEMTDMKULVHRVKNLSKRAIGLVGVKY 300
 QY 299 DHTDAVKSINELINAGIHTRRKQVSYISDSETEAEQTAALKNVDAILVPGGEGEV 357
 DB 301 SIQDAVLSVAEALRHAGYDHAIEIHWIDSEKVKENVAEIMKVDGILVGGGDBAI 360
 QY 358 EKKISTVFARENKTPYIGICGMSAIVFAPRVVLEGASHTFELKSPHPVIGLITE 417
 DB 361 EKKIAIVFARVKNKPYFYGICGMSAIVFAPRVVLEGASHTFELKSPHPVIGLITE 420
 QY 418 WMDPAGELVTRDSDLGITMRLGAQKCRLLKADSLAPOLYQDVITERRHRYEFENQYL 477
 DB 421 -----QNTIENMGITRLGILYPARIKOGTGAEAAGTTLVEERHRRHRYEFENQYL 470
 QY 478 KQLEAAGKTFSGSKIDGRUVBIIIEPPEPFLACQFHPFTSTPKNHALLPSCVYEAANK 537
 DB 471 EGMERAAQIVSATSPDGLVAVELIDHPVFAQYHPEIFSRPRLPQSLFADFVGAMLK 530
 QY 538 HK 539
 DB 531 NK 532
 RESULT 11
 AAU36779
 ID AAU36779 standard; Protein; 536 AA.
 XX AAU36779;
 DT 14-FEB-2002 (first entry)
 DE Staphylococcus aureus cellular proliferation protein #949.
 XX
 XX Antisense; prokaryotic cellular proliferation protein;
 KM antibiotic; antibacterial; drug design.
 XX Staphylococcus aureus.
 OS
 XX WO200170955-A2.
 PN 27-SEP-2001.
 XX
 PD 21-MAR-2001; 2001MO-US09180.
 XX
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR MPI; 2001-611495/70.
 DR N-PSDB; AAS54638.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 12372; 511bp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 536 AA;

Query Match 56.5%; Score 1576; DB 22; Length 536;
 Best Local Similarity 55.6%; Pred. No. 2.2e-136;
 Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

1 MTKFIFITGGVSSLSGKGIASSLAAILBDRGLKVTITLDPYINVDPTGMSPPQHGVEF 60
 1 MTKFIFITGGVSSLSGKGIASSLAAILBDRGLKVTITLDPYINVDPTGMSPPQHGVEF 60
 61 VTDDGAEITDLDGHERFLKTTMTKKNFTTGOVVEQVLRNKRKDYIGATVQVPHITD 120
 61 VTDDGAEITDLDGHERFLKTTMTKKNFTTGOVVEQVLRNKRKDYIGATVQVPHITD 120
 121 EIKRRVYESAE--GKVALIVEGTVGDIESTPLETTIQMGVEIGRDAALPHILTVPY 178
 121 EIKRRVYESAE--GKVALIVEGTVGDIESTPLETTIQMGVEIGRDAALPHILTVPY 178
 121 EIKRRVYESAE--GKVALIVEGTVGDIESTPLETTIQMGVEIGRDAALPHILTVPY 178
 121 EIKRRVYESAE--GKVALIVEGTVGDIESTPLETTIQMGVEIGRDAALPHILTVPY 178
 179 IKSAGELTKPTQHSVKELENTIGIQPDILICSEQPIPAESERRKIALFTVAEKAVISAI 238
 179 IKSAGELTKPTQHSVKELENTIGIQPDILICSEQPIPAESERRKIALFTVAEKAVISAI 238
 181 IKSAGEMKTKPTQHSVKELENTIGIQPDILIVTEVMTQDMDKIALPCDINKESITBOR 240
 181 IKSAGEMKTKPTQHSVKELENTIGIQPDILIVTEVMTQDMDKIALPCDINKESITBOR 240
 239 DADTTRIFILLRBCGLDVLVDQRLDVP--AADLSAMKVVNDGLTHPTDVSIAIVKY 297
 239 DADTTRIFILLRBCGLDVLVDQRLDVP--AADLSAMKVVNDGLTHPTDVSIAIVKY 297
 241 DADSLVEIPIQLSQNMDDIVIKRLQINKYETQDLEMKQLDIVNNLDGKITIGIVKY 300
 241 DADSLVEIPIQLSQNMDDIVIKRLQINKYETQDLEMKQLDIVNNLDGKITIGIVKY 300
 298 VDHDTAYKSLNREALIHAGIHTHRKVOISYIDSEFTIEAGTAK--IKNVDAIIVPGGGERG 356
 298 VDHDTAYKSLNREALIHAGIHTHRKVOISYIDSEFTIEAGTAK--IKNVDAIIVPGGGERG 356
 301 VSLQDAVLSVESLKRKAGYPPAKADIRKIDSEVTDENAAETLADVDGILVPGGGERG 360
 301 VSLQDAVLSVESLKRKAGYPPAKADIRKIDSEVTDENAAETLADVDGILVPGGGERG 360
 357 VEGKISTVFARENKIPLYGICLQMSAVIEFARNVVGEGASTTELPKSPHVIIGLT 416
 357 VEGKISTVFARENKIPLYGICLQMSAVIEFARNVVGEGASTTELPKSPHVIIGLT 416
 361 SEGRKIAIKYARENNVPFGICLQMLATVEFSRNVLGEGASALDAPATPYITDILP 420
 361 SEGRKIAIKYARENNVPFGICLQMLATVEFSRNVLGEGASALDAPATPYITDILP 420
 417 EMMDZAGELVTDDEDDLGTRVRLGOKRKLKADSLAPQLQKDVITTEHRRRYERNQY 476
 417 EMMDZAGELVTDDEDDLGTRVRLGOKRKLKADSLAPQLQKDVITTEHRRRYERNQY 476
 421 EQQD-----IDLGGLTLRLGLYPCSIKSGTLAGQVYGAALIEEHRHRRRYERNQY 470
 421 EQQD-----IDLGGLTLRLGLYPCSIKSGTLAGQVYGAALIEEHRHRRRYERNQY 470
 477 LKQLEAAGKPKSGSLDGLRLEVEILPEHPMFLACFHPSEFTSPENGHALFSGVEKAA 536
 477 LKQLEAAGKPKSGSLDGLRLEVEILPEHPMFLACFHPSEFTSPENGHALFSGVEKAA 536
 471 RBLQLEANGVITSTSPDGLRLEVEILPTNDFFLACQFHHBFLSRPHRPIKRSFIMBL 530
 471 RBLQLEANGVITSTSPDGLRLEVEILPTNDFFLACQFHHBFLSRPHRPIKRSFIMBL 530
 QY 537 KHK 539
 Db 531 KYQ 533

RESULT 12
 AAU37385
 AAU37385 standard; Protein; 536 AA.
 XX AAU37385,
 AC

XX 14-FEB-2002 (first entry)
 DT Staphylococcus aureus cellular proliferation protein #1555.
 DE Staphylococcus aureus cellular proliferation protein;
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX Staphylococcus aureus.
 OS
 PN WO200170955-A2.
 XX 27-SEP-2001.
 XX 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 XX 23-MAY-2000; 2000US-206848P.
 XX 26-MAY-2000; 2000US-207272P.
 XX 23-OCT-2000; 2000US-242578P.
 XX 27-NOV-2000; 2000US-253625P.
 XX 22-DEC-2000; 2000US-257931P.
 XX 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI WPI, 2001-611495/70.
 DR N-PSDB; AAS55244.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX
 XX Example 3; Seq ID No 12978; 511pp; English.
 RS
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 536 AA;

Query Match 56.5%; Score 1576; DB 22; Length 536;
 Best Local Similarity 55.6%; Pred. No. 2.2e-136;
 Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

1 MTKFIFITGGVSSLSGKGIASSLAAILBDRGLKVTITLDPYINVDPTGMSPPQHGVEF 60
 1 MTKFIFITGGVSSLSGKGIASSLAAILBDRGLKVTITLDPYINVDPTGMSPPQHGVEF 60
 61 VTDDGAEITDLDGHERFLKTTMTKKNFTTGOVVEQVLRNKRKDYIGATVQVPHITD 120
 61 VTDDGAEITDLDGHERFLKTTMTKKNFTTGOVVEQVLRNKRKDYIGATVQVPHITD 120
 121 EIKRRVYESAE--GKVALIVEGTVGDIESTPLETTIQMGVEIGRDAALPHILTVPY 178
 121 EIKRRVYESAE--GKVALIVEGTVGDIESTPLETTIQMGVEIGRDAALPHILTVPY 178

Db 121 EIKERLLAGESTNADVITEIGTGTGDIISLPPIAIRQIRSDIGRBNWVYHCTLLPY 180
 QY 179 IKSAGELTKPTQHSVKELRTIGIOPDILICRSEQPIPASERKIALFTNVAEKATSAI 238
 Db 181 IKAAGEMTKPTQHSVKELRTIGIOPDILIVRTEYENTODLKXIALPCDINKSEVIECR 240
 QY 239 DADITIRIPLLRBOGLDVLVDQLRDVP-AADLSAMEKXVGLTHPTDEVSAIATVGY 297
 Db 241 DADSLVEIPQLSQNMDDIVIKRLOLNKAYETQLDEWQQLDITVNNLQKITIGLVGYX 300
 QY 298 VDHDTAYKSLNEMLIHAGITRKAQVSIYDSTIEAGNKA-LKXNDAILVPGGGEKG 356
 Db 301 VSLQDAVLSVSESLKAGYFPAKDIDIRWIDSEBVDENAAETLADVDGILVPGGGEK 360
 QY 357 VEGKISTVRPARKNKIPYLGICLQMOASAVIEPARNVGLGAGHSTFELPKSPHVPGLIT 416
 Db 361 SEBKISAIKARAKNNVPFGICLQMOATVPSRNVLGEGHSAALDPAITPPIIDILP 420
 QY 417 EMNDEAGELVTRDESDLDGTMRLGAOKCRKLKADSLAFOLYOKOVITERRHRYEFNNQY 476
 Db 421 EOKD-----IEDLGITRLGLYPCSIKEGTILAOVYGKALEEERHRYEFNNQY 470
 QY 477 LKQLEAAGMKFSKSLDGRIVEIIELEPHWPLACQHPPEFTSPENGHALFSGFTEAAA 536
 Db 471 REOLESNAMVSGTSPDGRIVEIIEIPNDFFIACQFHPPEFTSPRPHPIFKSFEAL 530
 QY 537 KHK 539
 Db 531 NYQ 533

RESULT 13

AAG81965 standard; Protein: 535 AA.
 ID AAG81965 standard; Protein: 535 AA.

AC AAG81965;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1024.
 XX
 KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
 XX vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN MO200134809-A2.
 XX
 PD 17-MAY-2001.
 PF 09-NOV-2000; 2000MO-US30782.
 PR 09-NOV-1999; 99US-0164258.
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI kimmerly WJ;
 XX
 DR WPI, 2001-316495/33.
 DR N-PSDB; AAH52815.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 18; Page 297-298; 2189pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81965 to AAG81120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 535 AA;

Query Match 56.5%; Score 1575; DB 22; Length 535;
 Best Local Similarity 55.6%; Pred. No. 2.7e-136;
 Matches 302; Conservative 95; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKFFITGVVSSLSKGIASLSLAALIEDRGLKVTIKLPYINVDPTMSPFQHGVEF 60
 Db 1 MTKFFITGVVSSLSKGIASLSLAALIEDRGLKVTIKLPYINVDPTMSPFQHGVEF 60
 QY 61 VTEDGAETDLDLGHYERFLKTMTRKNNFTGQVYBOYLNERKGDYLGATVQVPIHTD 120
 Db 61 VTEDGAETDLDLGHYERFLKTMTRKNNFTGQVYBOYLNERKGDYLGATVQVPIHTD 120
 QY 121 EIKRRVYSAE--GNDVALIEVGGTVGDIEBFLFETIRQMGVLEGRLAFIHLTVLPY 178
 Db 121 EIKERLLAGESTNADVITEIGTGTGDIISLPPIAIRQIRSDIGRBNWVYHCTLLPY 180
 QY 179 IKSAGELTKPTQHSVKELRTIGIOPDILICRSEQPIPASERKIALFTNVAEKATSAI 238
 Db 181 IKAAGEMTKPTQHSVKELRTIGIOPDILIVRTEYENTODLKXIALPCDINKSEVIECR 240
 QY 239 DADITIRIPLLRBOGLDVLVDQLRDVP-AADLSAMEKXVGLTHPTDEVSAIATVGY 297
 Db 241 DADSLVEIPQLSQNMDDIVIKRLOLNKAYETQLDEWQQLDITVNNLQKITIGLVGYX 300
 QY 298 VDHDTAYKSLNEMLIHAGITRKAQVSIYDSTIEAGNKA-LKXNDAILVPGGGEKG 356
 Db 301 VSLQDAVLSVSESLKAGYFPAKDIDIRWIDSEBVDENAAETLADVDGILVPGGGEK 360
 QY 357 VEGKISTVRPARKNKIPYLGICLQMOASAVIEPARNVGLGAGHSTFELPKSPHVPGLIT 416
 Db 361 SEBKISAIKARAKNNVPFGICLQMOATVPSRNVLGEGHSAALDPAITPPIIDILP 420
 QY 417 EMNDEAGELVTRDESDLDGTMRLGAOKCRKLKADSLAFOLYOKOVITERRHRYEFNNQY 476
 Db 421 EOKD-----IEDLGITRLGLYPCSIKEGTILAOVYGKALEEERHRYEFNNQY 470
 QY 477 LKQLEAAGMKFSKSLDGRIVEIIELEPHWPLACQHPPEFTSPENGHALFSGFTEAAA 536
 Db 471 REOLESNAMVSGTSPDGRIVEIIEIPNDFFIACQFHPPEFTSPRPHPIFKSFEAL 530
 QY 537 KHK 539
 Db 531 NYQ 533

RESULT 14

AAB39346 standard; Protein: 561 AA.
 ID AAB39346 standard; Protein: 561 AA.

AC AAB39346;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4191.
 XX
 KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KM antibacterial; gene therapy.

Query Match 56.3%; Score 1569; DB 22; Length 536;
Best Local Similarity 55.4%; Pred. No. 9,7e-136;
Matches 301; Conservative 94; Mismatches 134; Indels 14; Gaps 4;

```
QY 1 MTKYIFITGGVSSLSGKGIASSLAAILLEDRLKVTITTKLDPIYINVDPGTMSPPQHGEEV 60
Db 1 MTKYIFITGGVSSLSGKGIASSLAAILLEDRLKVTITTKLDPIYINVDPGTMSPPQHGEEV 60
QY 61 VTEDGATDLDLIGYERFLTKMTKKNFTTGGVYBOYLNERKGDYLGATVQVPIPHITD 120
Db 61 VTEDGATDLDLIGYERFLTKMTKKNFTTGGVYBOYLNERKGDYLGATVQVPIPHITD 120
QY 121 EIKRVEESAE--GKDYALIEVGSTVGDIISLPLEFTIRQMGVELGRDRALFTHLTVPY 178
Db 121 EIKRVEESAE--GKDYALIEVGSTVGDIISLPLEFTIRQMGVELGRDRALFTHLTVPY 178
QY 121 EIKRVEESAE--GKDYALIEVGSTVGDIISLPLEFTIRQMGVELGRDRALFTHLTVPY 180
Db 121 EIKRVEESAE--GKDYALIEVGSTVGDIISLPLEFTIRQMGVELGRDRALFTHLTVPY 180
QY 179 IKSGLKTKPTQHSVVELTGTIGIOPDILICRSEQPIPASERKIALFTVAEKAVISAI 238
Db 179 IKSGLKTKPTQHSVVELTGTIGIOPDILICRSEQPIPASERKIALFTVAEKAVISAI 238
QY 181 LKAGEMKTPQHSVVELTGTIGIOPDILICRSEQPIPASERKIALFTVAEKAVISAI 240
Db 181 LKAGEMKTPQHSVVELTGTIGIOPDILICRSEQPIPASERKIALFTVAEKAVISAI 240
QY 239 DADTYIRIPLLRBOGLDLVDQLRLDVPADISAMERKVVDCGLTPDEVSAIVGKXY 298
Db 239 DADTYIRIPLLRBOGLDLVDQLRLDVPADISAMERKVVDCGLTPDEVSAIVGKXY 298
QY 241 DVEETLYSIPALQANNDQIVCDHKLIDAPADMTERRALBERKYNLKKTKRIALVGYV 300
Db 241 DVEETLYSIPALQANNDQIVCDHKLIDAPADMTERRALBERKYNLKKTKRIALVGYV 300
QY 299 DHTDAYKSLNBAIIHAGIHTRHVKQISYIDSETTEABGTA-KLKNVDAILVPGGFERGV 357
Db 299 DHTDAYKSLNBAIIHAGIHTRHVKQISYIDSETTEABGTA-KLKNVDAILVPGGFERGV 357
QY 301 ELPDAYSIVAEALKHAGFPDSDIEIDWVDSQELTYAENVALERIGSADGILVPGGFERGI 360
Db 301 ELPDAYSIVAEALKHAGFPDSDIEIDWVDSQELTYAENVALERIGSADGILVPGGFERGI 360
QY 358 ESKISTVRFARBNKIPIPLGICLGMQSAVIEFARNVVGLEBAGHSTEFLEKSPHVLGLITE 417
Db 358 ESKISTVRFARBNKIPIPLGICLGMQSAVIEFARNVVGLEBAGHSTEFLEKSPHVLGLITE 417
QY 361 ESKIEAIRFARNVDVPLGICLGMQSAVIEFARNVVGLEBAGHSTEFLEKSPHVLGLITE 420
Db 361 ESKIEAIRFARNVDVPLGICLGMQSAVIEFARNVVGLEBAGHSTEFLEKSPHVLGLITE 420
QY 418 WMDAGEIIVTRDSDLDGTRGLAOKRLKADSLAFOLY-QKQVITERRHRYEENNOY 476
Db 418 WMDAGEIIVTRDSDLDGTRGLAOKRLKADSLAFOLY-QKQVITERRHRYEENNOY 476
QY 421 -----QENIENLGGTLRLGLVPCCLKKGTAAAGNEDVQERHRYEENNOY 470
Db 421 -----QENIENLGGTLRLGLVPCCLKKGTAAAGNEDVQERHRYEENNOY 470
QY 477 LKQLEAAGKPFSGKSLDGLVETIELEPHEPFLACQHPHPTSTPRNGHALFSGFEEA 536
Db 477 LKQLEAAGKPFSGKSLDGLVETIELEPHEPFLACQHPHPTSTPRNGHALFSGFEEA 536
QY 471 RQLEENGIVSGVSPDNLVEIWEIPEKQFVACQFHPHLSRPNRPORLKGFVGAAL 530
Db 471 RQLEENGIVSGVSPDNLVEIWEIPEKQFVACQFHPHLSRPNRPORLKGFVGAAL 530
QY 537 KHK 539
Db 531 ANK 533
```

Search completed: January 29, 2004, 15:49:45
Job time : 51.7262 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 ; Search time 17.0734 Seconds
(without alignments)
1348.130 Million cell updates/sec

Title: US-09-941-947a-16

Perfect score: 2788

Sequence: 1 MKTFIPITGVSSLSKQIA.....HALFSGVEAAKHKTQGT 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2004	71.9	547	4	US-09-252-991A-21120
2	1825.5	65.5	567	4	US-09-328-352-7224
3	1575	56.5	561	4	US-09-134-001C-4191
4	1355.5	48.6	537	4	US-09-198-452A-246
5	1038	37.3	340	4	US-09-107-532A-6925
6	544.5	19.5	226	4	US-09-107-532A-4108
7	220.5	7.9	373	4	US-09-252-991A-25369
8	123.5	4.4	830	4	US-09-107-532A-5709
9	120	4.3	401	4	US-09-252-991A-27277
10	111.5	4.0	552	1	US-08-116-098-2
11	111.5	4.0	552	3	US-08-687-590-32
12	111.5	4.0	619	3	US-09-066-046-2
13	109.5	3.9	1407	4	US-09-328-352-7885
14	109.5	3.9	809	4	US-09-252-991A-3742
15	107.5	3.9	1700	4	US-09-252-991A-21763
16	107	3.8	773	4	US-09-252-991A-31403
17	106	3.8	171	2	US-08-560-098A-53
18	105.5	3.8	887	4	US-09-252-991A-16679
19	105	3.8	175	4	US-09-134-001C-4747
20	103.5	3.7	573	4	US-09-328-352-6420
21	103	3.7	483	4	US-08-887-534A-51
22	103	3.7	483	4	US-09-527-431-51
23	103	3.7	804	4	US-09-134-001C-5218
24	102.5	3.7	644	4	US-09-252-991A-25084
25	102.5	3.7	683	4	US-09-252-991A-32144
26	102	3.7	366	4	US-09-134-001C-4799
27	102	3.7	783	6	5231168-2

28	102	3.7	1027	3	US-08-446-137B-2	Sequence 2, Appl
29	101.5	3.6	458	4	US-09-198-452A-610	Sequence 610, App
30	101	3.6	236	3	US-08-935-263-2	Sequence 2, Appl
31	101	3.6	236	4	US-09-594-185-2	Sequence 2, Appl
32	101	3.6	775	1	US-07-603-133B-13	Sequence 13, Appl
33	100.5	3.6	582	4	US-09-252-991A-26782	Sequence 26782, A
34	100.5	3.6	1151	3	US-08-840-006-6	Sequence 6, Appl
35	100.5	3.6	1200	3	US-08-840-006-5	Sequence 5, Appl
36	100.5	3.6	3724	2	US-08-804-227C-10	Sequence 10, Appl
37	100.5	3.6	3724	2	US-08-804-198-4	Sequence 4, Appl
38	100	3.6	1044	4	US-09-252-991A-24495	Sequence 24495, A
39	100	3.6	1802	4	US-09-322-478-18	Sequence 18, Appl
40	99.5	3.6	524	4	US-09-252-991A-27783	Sequence 27783, A
41	99.5	3.6	551	4	US-09-252-991A-24209	Sequence 24209, A
42	99.5	3.6	858	4	US-09-255-828-22	Sequence 22, Appl
43	99.5	3.6	858	4	US-09-255-828-29	Sequence 29, Appl
44	99.5	3.6	921	4	US-09-252-991A-32974	Sequence 32974, A
45	99.5	3.6	1169	4	US-09-255-828-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-252-991A-21120
; Sequence 21120, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21120
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21120

Query Match	71.9%	Score 2004	DB 4	Length 547
Best Local Similarity	71.2%	Pred. No. 2.4e-188		
Matches 383	Conservative 63	Mismatches 92	Indels 0	Gaps 0
QY	1	MTKTFIPITGVSSLSKQIAASSIAAIIEDGKLTITKLDPIYINVDGTMSPFGHGEVF	60	
DB	6	MTRYIFVGVGVSSLSKQIASASIAAIIEDGKLTITKLDPIYINVDGTMSPFGHGEVF	65	
QY	61	VTEDAEVDLDIYHREFFLTITTKNNFTTGOVYEOYLNERKGDYVATQVPIHTD	120	
DB	66	VYQDAEVDLDIYHREFFLTITTKNNFTTGOVYEOYLNERKGDYVATQVPIHTD	125	
QY	121	EIKRRVYASAGKQVALIEVGCTVGDHSLEPFLITRQMGVELGRALPHLTLPYIK	180	
DB	126	EIKRRVYASAGKQVALIEVGCTVGDHSLEPFLITRQMGVELGRALPHLTLPYIK	185	
QY	181	SAGELKTPGTHSVYETRTIGIOPDILICRSBOPTPASERRKIALFTNVAEKAVISADA	240	
DB	186	TAGELKTPGTHSVYETRTIGIOPDILICRSBOPTPASERRKIALFTNVAEKAVISADA	245	
QY	241	DTIYRIPILREGGDLDLVVDQLRDVPAADLSAMERKVDGLTPTDEVSIALVGVYDH	300	
DB	246	DTIYRIPILREGGDLDLVVDQLRDVPAADLSAMERKVDGLTPTDEVSIALVGVYDH	305	
QY	301	TDAYKSLNEALIHGTHRHVQVSYDSTFTTEAGTAKLQNVALLVPGFGSGGVGCK	360	
DB	306	LDAYKSLNEALIHGTHRHVQVSYDSTFTTEAGTAKLQNVALLVPGFGSGGVGCK	365	


```

QY 361 ISTVFARENKIPYLIGICLQWQSAVIEFARNVVLGASHTEFLPKSPHPVIGLITWMD 420
DB 366 ISTVFARENKIPYLIGICLQWQSAVIEFARNVVLGASHTEFLPKSPHPVIGLITWMD 425
QY 421 EAGELVTRDSDSLGGMTRGAGQCRKADSLAQLYQKVITERRHRRHFNNOYLKOL 480
DB 426 ATGATIRTESSDSLGGMTRGAGQCRKADSLAQLYQKVITERRHRRHFNNOYLKOL 485
QY 481 EAAQKESGKSLDRLVAILIELEPHWFLACQFPEPTSTPRNGHALFSGFEAAAH 538
DB 486 EQAGLKISGSGDGLAVVEVSEPHWFLACQFPEPTSTPRNGHALFSGFEAAAH 543

```

RESULT 2

```

US-09-328-352-7224
; Sequence 7224, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Bretton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7224
; LENGTH: 567
; TYPE: PRN
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7224

```

```

Query Match 65.5%; Score 1825.5; DB 4; Length 567;
Best Local Similarity 64.6%; Pred. No. 9,3e-171;
Matches 351; Conservative 81; Mismatches 106; Indels 5; Gaps 3;

```

```

QY 1 MTKFIFITGVVSSIGKGIKAAASLAALIEDRGKTTIKDPIYINVDPGTMSFPGHGVF 60
DB 23 MTKFIFITGVVSSIGKGIKAAASLAALIEDRGKTTIKDPIYINVDPGTMSFPGHGVF 82
QY 61 VTDDGAEFTDLDGHERFELKTTMTKKNFTTGVVEQVLRNKKEDYLGATVQVPHITD 119
DB 83 VTDDGAEFTDLDGHERFELKTTMTKKNFTTGVVEQVLRNKKEDYLGATVQVPHITD 142
QY 120 DEIKRKYVSAEGDVALIEVGGTVGDIESLPLETRKQVGLGDRBALFHLTLVPI 179
DB 143 DTKERVLRAEGDVAIVIGTVGDIESLPLETRKQVGLGDRBALFHLTLVPI 202
QY 180 KSAEELKTKPTQHSVKELRTIGIQPDILICRSEOPIPASERRKIALFTVAERKVAISAD 239
DB 203 KSAEELKTKPTQHSVKELRTIGIQPDILICRSEOPIPASERRKIALFTVAERKVAISAD 262
QY 240 ADITIRPLILREOGDLDLVVDQLR-DVPAADLSAMEKVVQGLTHPTDEVSIAIVGKY 298
DB 263 AKTIYQIPRGFTYQNVDDLICERFGFTDLEADLTMDNVVLEALNPEYTVRAVMGKV 322
QY 299 DHTDAYKSLNEALIHAGIHTRHVKQISYIDSEFTEAEGTAKLVNDAIIVPGFGRGYE 358
DB 323 ELDDAYKSVNEALIHAGIHTRHVKQISYIDSEFTEAEGTAKLVNDAIIVPGFGRGYE 382
QY 359 GKSTVFARENKIPYLIGICLQWQSAVIEFARNVVLGASHTEFLPKSPHPVIGLITW 418
DB 383 GKSTVFARENKIPYLIGICLQWQSAVIEFARNVVLGASHTEFLPKSPHPVIGLITW 442
QY 419 MDAGELVTRDSDSLGGMTRGAGQCRKADSLAQLYQKVITERRHRRHFNNOYLK 478
DB 443 LDRGELVTRDSDSLGGMTRGAGQCRKADSLAQLYQKVITERRHRRHFNNOYLK 502
QY 479 QLEAAGKESGKSLDRLVAILIELEPHWFLACQFPEPTSTPRNGHALFSGFEAAAH 538
DB 503 ALEQAGKESGKSLDRLVAILIELEPHWFLACQFPEPTSTPRNGHALFSGFEAAAH 559
QY 539 KTG 541

```

```

DB 560 KTG 562

```

RESULT 3

```

US-09-134-001C-4191
; Sequence 4191, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4191
; LENGTH: 561
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4191

```

```

Query Match 56.5%; Score 1575; DB 4; Length 561;
Best Local Similarity 55.6%; Pred. No. 4e-146;
Matches 302; Conservative 95; Mismatches 132; Indels 14; Gaps 4;

```

```

QY 1 MTKFIFITGVVSSIGKGIKAAASLAALIEDRGKTTIKDPIYINVDPGTMSFPGHGVF 60
DB 27 MTKFIFITGVVSSIGKGIKAAASLAALIEDRGKTTIKDPIYINVDPGTMSFPGHGVF 86
QY 61 VTDDGAEFTDLDGHERFELKTTMTKKNFTTGVVEQVLRNKKEDYLGATVQVPHITD 120
DB 87 VTDDGAEFTDLDGHERFELKTTMTKKNFTTGVVEQVLRNKKEDYLGATVQVPHITD 146
QY 121 EIKRKYVSAE--GQVALLIEVGGTVGDIESLPLETRKQVGLGDRBALFHLTLVPI 178
DB 147 EIKRKYVSAE--GQVALLIEVGGTVGDIESLPLETRKQVGLGDRBALFHLTLVPI 206
QY 179 ISAGELKTKPTQHSVKELRTIGIQPDILICRSEOPIPASERRKIALFTVAERKVAISAI 238
DB 207 ISAGELKTKPTQHSVKELRTIGIQPDILICRSEOPIPASERRKIALFTVAERKVAISAI 266
QY 239 DADITIRPLILREOGDLDLVVDQLR-DVPAADLSAMEKVVQGLTHPTDEVSIAIVGKY 297
DB 267 DADITIRPLILREOGDLDLVVDQLR-DVPAADLSAMEKVVQGLTHPTDEVSIAIVGKY 326
QY 298 VHTDAYKSLNEALIHAGIHTRHVKQISYIDSEFTEAEGTAKLVNDAIIVPGFGRGYE 356
DB 327 VHTDAYKSLNEALIHAGIHTRHVKQISYIDSEFTEAEGTAKLVNDAIIVPGFGRGYE 386
QY 357 VEGKISTVFARENKIPYLIGICLQWQSAVIEFARNVVLGASHTEFLPKSPHPVIGLIT 416
DB 387 VEGKISTVFARENKIPYLIGICLQWQSAVIEFARNVVLGASHTEFLPKSPHPVIGLIT 446
QY 417 EMDAGELVTRDSDSLGGMTRGAGQCRKADSLAQLYQKVITERRHRRHFNNOYLK 476
DB 447 EMDAGELVTRDSDSLGGMTRGAGQCRKADSLAQLYQKVITERRHRRHFNNOYLK 506
QY 477 LKQLEAAGKESGKSLDRLVAILIELEPHWFLACQFPEPTSTPRNGHALFSGFEAAAH 536
DB 507 LKQLEAAGKESGKSLDRLVAILIELEPHWFLACQFPEPTSTPRNGHALFSGFEAAAH 556
QY 537 KTG 539
DB 557 KTG 559

```

RESULT 4

```

US-09-198-452A-246
; Sequence 246, Application US/09198452A

```

Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Griffiths, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 TITLE OF INVENTION: and treatment of infection
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 CURRENT FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 246
 LENGTH: 537
 TYPE: PRT
 ORGANISM: Chlamydia pneumoniae
 US-09-198-452A-246

Query Match 48.6%; Score 1355.5; DB 4; Length 537;
 Best Local Similarity 50.5%; Pred. No. 1.5e-124;
 Matches 271; Conservative 95; Mismatches 154; Indels 17; Gaps 7;

QY 3 KFIITGVVSSLGKGIASLSAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEFY 62
 DB 4 KCIPIFGVSSSLGKGIASLSAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEFY 63
 QY 63 EDGAEITDLDGHERPLKTKTKKNNFTTGOYVBOYLAKERKGDYIGATVOYIPHTTDEI 122
 DB 64 DDGEITDLDGHERPLKTKTKKNNFTTGOYVBOYLAKERKGDYIGATVOYIPHTTDEI 123
 QY 123 KRAVESAE--GKVALIEVGTVGDIESLPLETIRONGVELGRDRALEFILTLPYIK 180
 DB 124 IQYILDAKESPDYIVIEIGTIGISLPLEIRORRYHSD--CINIMTIVPYIQ 182
 QY 181 SAGELTKPTQHSVKEIKRTIGIOPDILCRSEQPIPASRRKIALFTNAEKAIVSAIDA 240
 DB 183 AADVEKSKPTQHSVKEIKRTIGIOPDILCRSEQPIPASRRKIALFTNAEKAIVSAIDA 242
 QY 241 -DIIYIPLLRGGDLDLVVDQLRDVPAADLSAEKVVDDGTHRTDEVSIAVSKYVD 299
 DB 243 KHTIYEPMLAOEKLAFIAGEKLTATVPENDRRVAVNQLSDLPYKIGVAGKYQ 302
 QY 300 HTDVAKSLNEALIHAGIHTRHAKVQISYIDSEITAEAGTAKLKNVDAILVPGRGERGVEG 359
 DB 303 HRDAYSIFEALHTHALRLGHAAEIPIDAE--DENLIMELSCDCLVPGRGVAGWEG 360
 QY 360 KISTVPAERANKIPYIGICLGMOSAVIEPAKRVGLEGHSTRELPKSPHPIGLITTEM 419
 DB 361 KIAAAKFCRBOGIPYIGICLGMQVAVVEARVNLMDQANSLEMPENTPHPIYV---M 416
 QY 420 DEAGELVTRDEDEDLGLTRELGAOKCRKLKADSLAFOLY--OKDVIYTERHHRRYEFNNQYAK 478
 DB 417 EGQDPLVA-----TGTRELGAIPCLLKPSKAKHAYNESSLIQRHHRRYEAVNDYIQ 470
 QY 479 QLEAAGKTSKSLDLRLVETIELEPEHMFACOPPEFTSTPRNGHALFSGPVEAA 535
 DB 471 SLIEDHGLRIVGTCPFGCEIIEVSDHPMMIGVQFPRFVSXKLISSPHPIAIFIAEA 527

RESULT 5
 US-09-107-532A-6925
 Sequence 6925, Application US/09107532A
 Patent No. 6583275

GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD/ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 6925:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 340 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...340
 SEQUENCE DESCRIPTION: SEQ ID NO: 6925:
 US-09-107-532A-6925

Query Match 37.3%; Score 1039; DB 4; Length 340;
 Best Local Similarity 59.2%; Pred. No. 9.7e-94;
 Matches 197; Conservative 61; Mismatches 63; Indels 12; Gaps 2;

QY 1 MTKFIITGVVSSLGKGIASLSAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEFY 60
 DB 7 MTKYIFVGVVSSLGKGIASLSAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEFY 66
 QY 61 VTDEGAETDLDGHERPLKTKTKKNNFTTGOYVBOYLAKERKGDYIGATVOYIPHTTDEI 120
 DB 67 VTDBAEITDLDGHERPLKTKTKKNNFTTGOYVBOYLAKERKGDYIGATVOYIPHTTDEI 126
 QY 121 EIKRIVESAE--GKVALIEVGTVGDIESLPLETIRONGVELGRDRALEFILTLPYIK 178
 DB 127 EIKERIKMAAQTSDVITTEVGTVGDIESLPLETIRONGVELGRDRALEFILTLPYIK 186
 QY 179 IKSAGELTKPTQHSVKEIKRTIGIOPDILCRSEQPIPASRRKIALFTNAEKAIVSAIDA 238
 DB 187 LKAAGEMTKPTQHSVKEIKRTIGIOPDILCRSEQPIPASRRKIALFTNAEKAIVSAIDA 246
 QY 239 DADTIYRIPLLRGGDLDLVVDQLRDVPAADLSAEKVVDDGTHRTDEVSIAVSKYVD 298
 DB 247 DADTIYRIPLLRGGDLDLVVDQLRDVPAADLSAEKVVDDGTHRTDEVSIAVSKYVD 306
 QY 299 DHTDAKSLNEALIH-----AGIHTRHAK 321
 DB 307 ELDPAYISVEALKHLIVSPITTSRSISGRK 339

RESULT 6
 US-09-107-532A-4108
 Sequence 4108, Application US/09107532A
 Patent No. 6583275

GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310

```

CORRESPONDENCE ADDRESS:
ADDRESSES: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4108:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...226
SEQUENCE DESCRIPTION: SEQ ID NO: 4108:
US-09-107-532A-4108

Query Match      19.5%; Score 544.5; DB 4; Length 226;
Best Local Similarity 47.4%; Pred. No. 2.2e-45;
Matches 110; Conservative 38; Mismatches 71; Indels 13; Gaps 4;

QY 313 HAGIHTKAVQISYIDSEITEAGT-AKLNVDAILVPGSGRGVEGKISTVRARENK 371
DB 5 HSGFAYNSDEIDBIKAKQELTRENVERLKDADGILVPGSGRGVEGKISTVRARENK 64
QY 372 IPIGLICQKQSAVIEFARNVVGIGAHSTFELPKSPHVIIGLITMMDAGELVTRDED 431
DB 65 VPEGLICQKQSAVIEFARNVVGIGAHSTFELPKSPHVIIGLITMMDAGELVTRDED 114
QY 432 SDLGTRRLGAQKRLADSLAFOLY-QKDVITERHRYEPNNQYKOLEAAGMKESGK 490
DB 115 ENLGTRRLGAQKRLADSLAFOLY-QKDVITERHRYEPNNQYKOLEAAGMKESGK 174
QY 491 SLGRLVEILDEHPWFLACQHPFETSTPRNGHALFSGFVAA-AKENTQ 541
DB 175 SPNRLVEILDEHPWFLACQHPFETSTPRNGHALFSGFVAA-AKENTQ 226

RESULT 7
US-09-252-991A-25369
Sequence 25369, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A

```

```

CURRENT FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 25369
LENGTH: 373
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25369

Query Match      7.9%; Score 220.5; DB 4; Length 373;
Best Local Similarity 27.8%; Pred. No. 3.9e-13;
Matches 72; Conservative 38; Mismatches 124; Indels 25; Gaps 4;

QY 283 THPTDEVSTAIKXVDHTDAKSLNEALHAGIHTRKQVQISYIDSEITEAGTAKLN 342
DB 20 TEPAPRLRIALVGDHPHTAHRAIPLALRLAGEALGLEIAFDWLASDRLPAE--PALER 77
QY 343 VVALIPGSGRGVEGKISTVRARENKIPYLGLICQKQSAVIEFARNVVGIGAHSTF 402
DB 78 YGFWCVPQSPYRDADVRLTAHARGRRRPPGLTQPGQHTTLEFARNVVGIGAHSTF 137
QY 403 FLPSHPHVIIGLITMMDAGELVTRDESDIGGTRRLGAQKRLADSLAFOLYQKDV 462
DB 138 EHPHSDQAVIALPALLEAREDV-----RLFGRLALVAADWT 178
QY 463 TERHRRYEPNNQYKOLEAAGMKESGSLDGLVEIIEHPHWPFLACQHPFETSTPR 522
DB 179 EADYHCRYAIAPFAALITGALRASMSADG-AIRAVELQHPFVATLPQBRRAVAG 237
QY 523 NGHALFSGFVAAAKENTQ 541
DB 238 VAPLPKAFVENC---RTQ 253

RESULT 8
US-09-107-532A-5709
Sequence 5709, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A. Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

```

INFORMATION FOR SEQ ID NO: 5709:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 830 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (8) LOCATION 1...830
 SEQUENCE DESCRIPTION: SEQ ID NO: 5709:
 US-09-107-532A-5709

Query Match 4.4%; Score 123.5; DB 4; Length 830;
 Best Local Similarity 20.4%; Pred. No. 0.0051;
 Matches 115; Conservative 80; Mismatches 201; Indels 169; Gaps 27;
 15 LGKGIASSIAALLERGLKAVITIKDPIYINVDPGTMSFONGEV--FYTE-----DG 65
 258 MGLVAGTKYRGFEEDRMKV-----IDBIYN-----DQVILFLDELTLIGAG 303
 66 AETDLDGHERFLKTKTKKNNFTTGQV---YEQVLNKKRGDTLGTQVYIPIITDS 121
 304 AEGALDASN---ILKPAALAGELQITIGATTLDEYQYIKDSALERRFARICVDEPTPEE 360
 122 IK-----RREYSSAKGQVALIEGQVQGDIESLPFLFTIRQMYELGRDALFIHTL 175
 361 AEVILQGLRRYEHGVEITDEAVPAAYN--LSYNT--TSROL-----PDKAIDL----- 408
 176 VPIKAGELKTKPTQHSYKELRTIGIQPDILICRSEQPIPASERRKIALFTNVV----- 230
 409 --IDESAAKRLQOTDLTKS--TVIKLEIDELVQEKAAIQKOD-----PENAAQLRRQ 459
 231 EKAV-----ISALDDPTIRIPLLRBGLDDIVVDLRLVPADLSAMTKV----- 278
 460 EKALRKQLQKVSALBK-----QEGYSDRYER--DV--ATVSEMGVPLQGLE 506
 279 -----VDGLTPPTDEVSIAVGRVYDHTDAVKSINELIHAGIHRHKVQLSYIDS 329
 507 KXSERULBELBGLHS-----RVVGQDEAVAVASRAIRRA----- 541
 330 ETTIAGCTAKLKNVD-----AIIVPGGEGRCVCEKISTVPRANKIPIYIGICLQKMS 383
 542 -----RSGIKNDPDRPIGSPFMTGPTGVKTEIALALSEVMFGSDAL-----IRV 586
 384 AVIEFARNVAGLEGASTFELPKSPHPIVGLITEMVDBAGELVTRDESDLGSTWELGAQ 443
 587 DMSEFVEK-----YSTSLISSPPGYG-----YDEGQLTEKTRQKPYVYIILDEVE 634
 444 KCRILKADSLAFQYQKDVITERHRRHYEFNN-----QYLKQLEBAQKFSGKSL 492
 635 KAPDVFNILLQVLDGHLTDSGRKVDPRNTIMTMSIGATQIREEKNVGNVQDVTK 694
 493 DGLVLEILHEHPHFLACQFHPEP 517
 695 DHRAPQRIIE---EKQAFRPEP 715

RESULT 9
 US-09-252-991A-27277
 Sequence 27277 Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: MARC J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252.991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074.788
 PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094.190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 27277
 LENGTH: 401
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27277

Query Match 4.3%; Score 120; DB 4; Length 401;
 Best Local Similarity 20.7%; Pred. No. 0.0034;
 Matches 82; Conservative 53; Mismatches 153; Indels 108; Gaps 17;
 177 PIKSGELKTKPTQHSYKELRTIGIQPDILICRSEQPIPASERRKIALFTNVAKAVIS 236
 81 PHIGNTG---TTPBAENRWMAAG-----LIHDLPIASNNRKSQSLPYLKANGTVA 132
 237 AIDADTIRIPLLRBGLD-----LVVDQLRLDVPADLS 273
 133 IAGIDT--RRLTRILAEKSGONGCIIAGADATSERALBLARAFPGIKGNDLAKRVTAEY 191
 274 AMEKVVDGL---THTDVSIAIVGRVYDHTDAVK-----SLNBALIHAGIHRHKVOIS 325
 192 ERRSSVWNLSDSHP--RIP---AGELPYHVAVDYGVKAILRLVLARG-----CRLS 240
 336 YIDSETIABGTAKLKNVDAILVPGFGE--RGVEKISTVFARENKIPIYIGICLQKMS 383
 241 VVPAQTAPASVLA--LAPDGIFLSNGPDPPEPCVYAIQAIREFLPTERIPIVGI CLGHOLL 298
 384 AVIEFARNVAGLEGASTFELPKSPHPIVGLITEMVDBAGELVTRDESDLGSTWELGAQ 443
 299 ALASGAKTLKQGHGHGAN-----HPV-----QDLDS---GVVMTISQ 333
 444 KCRILKADSLAFQYQKDVITERHRRHYEFNNQYLKQLEBAQKFSGKSLDGLVLEIIEP 503
 334 NHGFVDSSTL-----PDILRATHKSLPDTIQIGIERTKVAFSPFG----- 375
 504 EHPWFLACQFHPEFTSTPRNGHALFSGFVEAAHK 539
 376 -----HPEASFGPHDVAFLFDRFISAMARR 401

RESULT 10
 US-08-116-098-2
 Sequence 2, Application US/08116098
 Patent No. 5428131
 GENERAL INFORMATION:
 APPLICANT: Trent, Jonathan D.
 APPLICANT: Horwich, Arthur L.
 TITLE OF INVENTION: Archaeobacterial Chaperonin-Mediated
 TITLE OF INVENTION: Protein Stabilization
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Kilpatrick & Cody
 STREET: 1100 Peachtree Street
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: US
 ZIP: 30309
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/116.098
 FILING DATE:
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/756627
 FILING DATE: 09-SEP-1993
 APPLICATION NUMBER: US 07/721974
 FILING DATE: 27-JUN-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/673158
FILING DATE: 18-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/261573
FILING DATE: 24-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Padst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: YU102
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Sulfolobus shibatae
US-08-116-098-2

Query Match 4.0%; Score 111.5; DB 1; Length 552;
Best Local Similarity 21.1%; Pred. No. 0.04;
Matches 129; Conservative 73; Mismatches 176; Indels 233; Gaps 31;

5 IFITGVVSSLGKIAASSLAAL-LBD-----RGL-KVTITLDPYINVDPGTMS 53
16 IILKSSSRITGKALANAAVAIAEALKSTYGRGDMKMFVDSL----- 62
54 FQHGKAVFTEDGAE-TDLDLGHYERFLKTYTKKNNFTTGQVYEGVLRNKKGDYLGAT 111
63 ---GDITITNDGATILDKMLOH-----PTGKLIVQIAKGQ----- 95
112 VQVIFHITDEIKRKYVESAGKQVALIEVGTVGDIESLPLETTIRQMGVELGRDALFI 171
96 -----DE-----ETADGTTAVILAGELAKKAEDILYKE----- 125
172 HLT-LVPYIKSAGELKTKPTQSHSVKELRTIGIOPDILICRSBOPIPASER---RKIALFT 227
126 HPTIIVSGYKKAELAKTIQ-----DI-----AQPVISINDTVLKRKVAL-F 166
228 NVAEKAVISALDADTYIRIPLLRKGLDLDVND-----QLRDVPAADLSAMB--KVV 279
167 SLGSKAVAGA-----REYLDLVVKAQAVALRGDKRYVDLDNVQIVKQH 212
280 DGLTHTDVSIAIVGYVDHTDAYKSLEALIHAGIHRKHYQVSIYDSEF----- 331
213 GGSINDTVLVGIVDKEVHHPMPRIENAKI-ALLDASLEVEKEPELDALIRINDPTQM 271
332 ---IEAEGTAKLVNDAIIVPGG---FGRGVE-----GXISTVFARENNIPYIG 376
272 HKFLKEEENILMEKDKIAATGANNVTCQKIDENVAQHIAKGLIARAKGSLDLEKTA 331
377 ICLGMO-----SAVIFPA--NVGLEGASHTFLPKSPHPVI-GLIT 416
332 RATGKAVISINIDELTSQDVGVAALVEERKVGEDKRVFVGAGN---PXSVALIRGGL 387
417 EWMDEAGELVTRDESDLGITVRLG-----AOKCRLA-----DSLAPOLY 457
388 RYVDEI-EKALDADLTADVTRDGRAVAGGAVETELKRLRKAPQVGGSEQLAIYAY 446
448 QKDV-----ITERHRYEFNNQYLKOLEAAGMKFSGKSLDGRIVE 498
447 ANAIGLILMIAENAGLDPIDKMQRSIHENETKMYGALN-----FTGN----- 492
499 IIEPEHPFL 509

DB 493 ----PEDMWKL 499

RESULT 11
US-08-687-590-32
Sequence 32, Application US/08687590
Patent No. 6255070
GENERAL INFORMATION:
APPLICANT: Willisson, Keith Robert
APPLICANT: Kubota, Hiroshi
APPLICANT: Ashworth, Alan
TITLE OF INVENTION: Folding Proteins
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,590
FILING DATE: 31-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/00192
FILING DATE: 31-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401791.0
FILING DATE: 31-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9418234.2
FILING DATE: 09-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin J.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 084619-000000US
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-687-590-32

Query Match 4.0%; Score 111.5; DB 3; Length 552;
Best Local Similarity 21.1%; Pred. No. 0.04;
Matches 129; Conservative 73; Mismatches 176; Indels 233; Gaps 31;

5 IFITGVVSSLGKIAASSLAAL-LBD-----RGL-KVTITLDPYINVDPGTMS 53
16 IILKSSSRITGKALANAAVAIAEALKSTYGRGDMKMFVDSL----- 62
54 FQHGKAVFTEDGAE-TDLDLGHYERFLKTYTKKNNFTTGQVYEGVLRNKKGDYLGAT 111
63 ---GDITITNDGATILDKMLOH-----PTGKLIVQIAKGQ----- 95
112 VQVIFHITDEIKRKYVESAGKQVALIEVGTVGDIESLPLETTIRQMGVELGRDALFI 171
96 -----DE-----ETADGTTAVILAGELAKKAEDILYKE----- 125
172 HLT-LVPYIKSAGELKTKPTQSHSVKELRTIGIOPDILICRSBOPIPASER---RKIALFT 227
126 HPTIIVSGYKKAELAKTIQ-----DI-----AQPVISINDTVLKRKVAL-F 166
228 NVAEKAVISALDADTYIRIPLLRKGLDLDVND-----QLRDVPAADLSAMB--KVV 279
167 SLGSKAVAGA-----REYLDLVVKAQAVALRGDKRYVDLDNVQIVKQH 212

QY 280 DGLTHPTDEVSIAIVCKYVDHTDAVKSLEALIHAGITRKRKVOISYIDSEF----- 331
 DB 213 GGSINDTOLVYGVIVDKEVHPKPKRIENAKI-ALDASLEVEKPELDAERINDPTQM 271
 QY 332 ---IHAEGPAKKNVDAILVPGG---FGERGVE-----GKISTVPARENKIPYLG 376
 DB 272 HKPHEBENILKERVKDLATGATNAVYTCCKGIDEVAGQHTLAKKGLAVARAKKSLBKL 331
 QY 377 ICIAGQ-----SAVIEFAR-----NVVGEAGHTEFLPKSPHPVI-GLIT 416
 DB 332 RATGGRVINSIDELNLSQDLGYAALVEBKVGEDKVVPEGAAN-----PKSVAILRGLE 387
 QY 417 EMMDEAGELVTEDESDLGSTWELG-----AOKCKLKA-----DSLAPQLY 457
 DB 388 RVVDET-ERALDRLAGTAVDIRDGGAVAGGAVELEIAKRLKRYAPOVGKQELAIAY 446
 QY 458 QKDV-----ITERHRYEFNNQYLKQLEAAGMKFSGKSLDGRIVE 498
 DB 447 ANAIEGLIMILANMGLDPIDKLMQRLSHENETKRYGLNL-----FTGN----- 492
 QY 499 IIELEPHWEL 509
 DB 493 ----PEDMWKL 499

RESULT 12
 US-09-066-046-2
 Sequence 2, Application US/09066046A
 Patent No. 6204252
 GENERAL INFORMATION:

APPLICANT: MURPHY, Cheryl

STOREY, James
 COUGHLIN, Richard T.
 BEUTZ, Gerald A.

TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC
 EHRILICHA AND METHODS OF USE

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESSER: HALE AND DORR LLP

STREET: 60 State Street
 CITY: Boston
 STATE: MA

COUNTRY: United States

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,046A
 FILING DATE: 24-Apr-1998
 CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: Superko, Colleen
 REGISTRATION NUMBER: 39,850
 REFERENCE/DOCKET NUMBER: 106,941.155

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 526-6900
 TELEFAX: (617) 526-5000

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 619 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-066-046-2

Query Match 4.0%; Score 111.5; DB 3; Length 619;
 Best Local Similarity 21.1%; Pred. No. 0.048;
 Matches 95; Conservative 62; Mismatches 178; Indels 115; Gaps 20.

QY 44 INVPGTMSPP-FOGEVFTEDGATDLDLGHYRFRFKTTTKKNTFTTGQVYQ----- 97
 DB 223 VQVESSTIVPESQGTDTVLEDTTETITVDEGCHFDIASGEHNNDLPAVLLDEADPTM 282
 QY 98 VLNRKRGDYLGAIVQVITPHITDEIKRKYVE-----SABEK-----DVALIE 139
 DB 283 LIANEE-----SKTIESMP--SDSELDNVQELGTLPLQGEHTVSEGTRESLPTDVSQDS 335
 QY 140 VG-----GTGDIRESLPLEET-IRQMGVEIGRDLPIHITLVPIYKASGLK 186
 DB 336 VGVSTDLAHSQEVETVSEVSTQDSLTNSIQDSVGVSTD--LHAKSKGVIEGG--- 390
 QY 187 TKPTQHSVK---ELRTGIQPDILICSEQPIPASERKIALFTNVAKAVISALDPTI 243
 DB 391 ---TQDSLSADPPIINTVSEESTDLEAHSQEVETVSEFTQDSLTNSIQDSVGVSTDLVH 447
 QY 244 YRIPLLRBOGLDLYVDQRLDY--PAADLSAMEKVYDGLTHPTD-----VSLAIVGK 296
 DB 448 SQEVEIVSEGTQDSLTNSIQDSVGVSTDLAHSQEVETVSEFTQDSLTNSIQDSVGV 507
 QY 297 YVDHTDAVKSLEALIHAGITRKRKVOISYIDSEFTIRABGTAKKNVDAILVPGGFERG 356
 DB 508 STD-LEVASGEVIEVSEGTQDSLTNSIQ--DSVGVSTDLAHSQGVITV-----SEGG 559
 QY 357 VBKISTVPARENKIPYLGICLQKQSVIFAPKNVGLSEANSTFELPKSPH----- 409
 DB 560 TQDSLSA-----DFPINTVSE--STDLEAHSQEVIESEV 593
 QY 410 ----FVIGLITBMWDBAGELVTRDESDL 434
 DB 594 STQAPSTGVETIRPVD-----RDSDDV 616

RESULT 13
 US-09-328-352-7885
 Sequence 7885, Application US/09328352
 Patent No. 6562958
 GENERAL INFORMATION:
 APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC99-03PA
 CURRENT APPLICATION NUMBER: US/09/328,352
 CURRENT FILING DATE: 1999-06-04
 NUMBER OF SEQ ID NOS: 8252
 SEQ ID NO 7885
 LENGTH: 1407
 TYPE: PRT
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-7885

Query Match 3.9%; Score 109.5; DB 4; Length 1407;
 Best Local Similarity 19.8%; Pred. No. 0.29;
 Matches 139; Conservative 91; Mismatches 223; Indels 249; Gaps 38.

QY 41 DRYINVDGTPSPF-----OHGEVFTEDGAT-----DLDLGH----- 74
 DB 157 ESYVYTDG-MTFPKYGLNDEEYFTLBEHGDFPAKKAARAVQDLKQIDLEABISR 215
 QY 75 -YERFLKTYMTKK-----NN-----F 89
 DB 216 LREEIPQTTSEKLLKAKSKRLKMEAFKDSNNKEMWNVLPVLPDLRLVPLEGRF 275
 QY 90 TT---GVYEVYL-----RNE-----RKGYL-GAT 111
 DB 276 ATSLDNLIRKVINRNNRLKLLDLAAPDIYIKNEKTLQSSVDALLDNKGRGRATISGN 335
 QY 112 VQVPHITDEIK--RKYVSAEGKV-----ALIEVGTVGDIE-SLPLETTRQMGVE 162

```

Db 336 KAPPLSLAMIKGKGRFRONLIGKRVDSGSVITVSTPLRLHQCGLP-----KMALE 390
QY 163 LGDRALFTHLTU-----VPYIKSAGELTKPTQHSVKELRTIGIQPDLICRSQPIPA 217
Db 391 LFPAP---FPAKLQASGQATTIKAAKQWERETPEVWDVLAIVIRQHFWMLRA-----PT 443
QY 218 SERKALTLTNAEKVAVISAIQADTYRIPLIREGLDLDVVDQLRLVP---AALISA 274
Db 444 LHLGLQAEPI-----LIEGKALRLHPLVCAAFNA-DPQDQWAVHVPILTLBAQLBA 495
QY 275 MEKVU-----DGLTHPTDEVSIAIVGKYVDHTDAYKSLNEALIIA----- 314
Db 496 RALMSTNNILSPANEPILIVSQD---VLGIYITRAVNAKGEWFAADTHEVNAAL 552
QY 315 ----GITHKXQVSIYIDSETTEABGTAKLKNVDAILVPG-----GGEKGVGKISTV 364
Db 553 ATQGVAVIHARVVR---VHQVTVINENGEBEQIIVDTTPGRCLLMEVVPBGLSTDMNL 609
QY 365 RFAREKRIPIYIGC---LQWASAVIEFARNVVLGEGAHSTE-----FLPKSPHVV 411
Db 610 EMTKKNISKLINSCTKRLGKQTVI-FADQLATLIGFRQATRSVSGVMEDMLIPETKTI 668
QY 412 I-----GITTEWMDAGELVTRDESDLGTMRLGAQKCRUKADSLAQ 455
Db 669 IDKATEVEBEIQOFRQGVLT---AGERYNKVVDI---WARTVDQAKAMMDNLSTY 719
QY 456 LYO-KVYTERHRRHREFFNOY-----LKOLEAAGMKFSGSLDGLRVEI-I 500
Db 720 LVNKKQGEDEKQ---SPNSIYMSDSGARSAQIRU---AGMRGLMAKPDGSIITETPI 774
QY 501 ELBEHPWFLACQHPPEFTSTPRNGHALFSGFVZAANKKXQG 542
Db 775 KANRREGTLVQY---FIST---HGARKGLADTALKTANSNG 809

```

```

RESULT 14
US-09-252-991A-32742
; Sequence 32742, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32742
; LENGTH: 809
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32742

```

```

Query Match 3.9%; Score 109; DB 4; Length 809;
Best Local Similarity 19.2%; Pred. No. 0.13;
Matches 126; Conservative 74; Mismatches 203; Indels 252; Gaps 31;

```

```

QY 42 PYINDRCMTSPFQHGIVFVTEGQ-----ETDLDLGHIERFLKTTMKNNFTTGOYV 95
Db 19 PGVAVAGAPPIVAHVAVOPTLDAVVAITQGPVDLWHLRQRKRPDRDHPQVNRORF 78
QY 96 -----EQLVNERKSG-DYLGATVQVAPHTIDEIKRNV-----YESABKDV 135
Db 79 GQRTDVLAAQALGEGAGAPFVGVDVDAHQALAEVLASVGNPFDVAAVADAVRBPV 138
QY 136 ALIE-----VGTGVD-----ISLPLETTRQWG---VELGRBALFIHTLVP 177
Db 139 AFABQPPAALAGEVABGLDAHRLAVVEDRHVLAQHRGLSGKQVAAIARQAV----- 190

```

```

QY 178 YTKSAGELTKPT-----QHSVKELRTIGIQ-PDIL----- 207
Db 191 -----GQYQVDPSSGGVQAPPAVQAQHNLDPVAQRAEQPLHQRLRQEPDLAVAPQDR 245
QY 208 -----ICRSEQPIPASERKIALF-TNVAEKA---VISAIDDTIYRIBLL 250
Db 246 VYRIVAIADGIRSLCQAQO-PSABRQPAIHSGVAGRGEQVAGAVDQ----- 295
QY 251 BEQGLDLDVVDQLRLVPADLS---AMEKYVDGLTHPTDEVSIAIVGKYVDHTDAYKSLN 308
Db 296 -----DQLFRGVVQALQADQDHYD-----ALAQQLAVAD----- 329
QY 309 EALIHAGIHTHKQVSIYIDSETTEABGTAKLKNVDAILVEGGEKGVGKISTVFEAR 368
Db 330 --LVH-----QVAAQKPRVADBSAQLBEPANGEGCFVARKV----- 365
QY 369 ENKIPYIGICLQWASAVIEFARNVVLGEGAHSTE-----HSTELPKSPHVV 413
Db 366 DQAP---VGVQPVAVELPEGLPBGAGLGAHAPADQILRPHQFARVQRDLHYIVG 420
QY 414 LITTEWMDAGELVTRDESD---LGTMRLGAQKCRUKADSLAQVQKDVITE----- 464
Db 421 AALQ-ADDAIDLVAAPGDQDPHLGAHPQLAGQCAVLAGQADVQKQDLALAEWALGX 479
QY 465 -----HRR---HRYEFNNQYL-----KOLEAAGMKF 487
Db 480 AGFRLGVAGADVYALALAEIGLEKLTHRVVVDHQVARNRHHLAPAGRGVSCSGLSD 539
QY 488 SKSLDGLVAVIILPEHPWFL---ACQFH--PEFTSTPR-----NGHAL 527
Db 540 GGRKGPGL-----PLPAAVPLNLGRSTSFRTPRAMAGARLQAIHPGLANTHAL 550

```

```

RESULT 15
US-09-252-991A-21763
; Sequence 21763, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21763
; LENGTH: 1700
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21763

```

```

Query Match 3.9%; Score 107.5; DB 4; Length 1700;
Best Local Similarity 22.0%; Pred. No. 0.64; Indels 97; Gaps 16;
Matches 84; Conservative 43; Mismatches 156;

```

```

QY 110 ATVCYPIPHITDEIKRKYVESABKQVALIEVGTVG---DIESLPLETTRQMGVELGR 165
Db 534 AVEVEVGR-----REVVPRAQGVADVMQVAAHQHVALAQRTVGCOPTLRLGGVDHRR 588
QY 166 DRALFIHTVPIYISABE-----LKTPTQHSVVELRTIGIQ-DI 206
Db 569 EDLAVHLDLPQHDVAVQRRHLRAQHTQLQVQLVVSQGVNHQVAFQRVAAQPIQDV 648
QY 207 LICR-----SEQPIPASERKIALFTNVAEKAVIS---AIDDTIYRIBLL---RE 252
Db 649 ALAGARHHRVADQPL-----LVEVVAEALASVAVHQAQAEQVAVREBELQFGERR 699
QY 253 QGLDLDVVDQLRLVPADLSAMEKYVDGLTHPTDEVSIAIVGKYVDHTDAYKSLNEALI 312

```



```

Db      700 VGLDQVIVAGARR-----GVAGHALDPGDARQATERRA 733
QY      313 HAGIHRHKVOISYIDSETEAEQT-----AKLKNVDALIVPGPG--ERGVE-GKIST 363
Db      734 ALILDVGRASQA--VDSHALALPCTGVAGALLADRDATATGAVGAGTERGAALAVIGE 791
QY      364 VRFARENKIPYLGLICMOSAVIEFARNVVG--LEGHSTETLPKSPHPVIGLITEMMDE 421
Db      792 ARAETEOAVVTLGOAENGQGNRYDLAPRGVGRQLRGQRHHRNPAAHRAVAHILAAADE 851
QY      422 ---AGELVTRDE---DSDLG 436
Db      852 GVAGAGDAVVDARAGIDRDVAG 873

```

Search completed: January 29, 2004, 15:57:07
 Job time : 18.0734 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 37.1346 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-16

Perfect score: 2788
Sequence: 1 MTKFIFITGVVSSIGKXIA.....HALFGFVEAAKKTQGA 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCT05_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09A_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2788	100.0	544	10	US-09-934-903-12
2	2788	100.0	544	10	US-09-934-903-12
3	2788	100.0	544	11	US-09-941-947a-16
4	2008	72.0	542	9	US-09-815-242-11933
5	1908.5	68.5	545	9	US-09-815-242-11933
6	1908.5	68.5	545	9	US-09-815-242-11933
7	1783	64.0	511	9	US-09-815-242-11933
8	1576	56.5	536	9	US-09-815-242-12372
9	1576	56.5	536	9	US-09-815-242-12372
10	1569	56.3	536	9	US-09-815-242-4984
11	1569	56.3	536	9	US-09-815-242-10727
12	1561	56.0	535	9	US-09-815-242-13376
13	1444.5	51.8	538	9	US-09-815-242-11320
14	1432.5	51.4	549	15	US-10-156-761-14031
15	1372	49.2	554	10	US-09-738-626-5066

16	1355.5	48.6	537	12	US-10-289-762-246	Sequence 246, App
17	1006.5	36.1	1003	12	US-10-161-051-91	Sequence 91, App
18	513	18.4	298	12	US-10-264-049-2881	Sequence 2881, App
19	353.5	12.7	199	15	US-10-106-698-4980	Sequence 4980, App
20	345.5	12.4	195	9	US-09-925-237-773	Sequence 773, App
21	327.5	11.7	94	12	US-09-864-408A-3914	Sequence 3914, App
22	245	8.8	59	12	US-09-864-408A-2350	Sequence 2350, App
23	204	7.3	135	12	US-10-264-049-4217	Sequence 4217, App
24	126.5	4.5	238	12	US-10-314-657-34	Sequence 34, App
25	126	4.5	374	12	US-10-369-493-10331	Sequence 10331, App
26	125.5	4.5	1390	12	US-10-369-493-11073	Sequence 11073, App
27	124	4.4	374	12	US-10-369-493-15844	Sequence 15844, App
28	124	4.4	374	12	US-10-369-493-16223	Sequence 16223, App
29	123.5	4.4	371	12	US-10-369-493-29	Sequence 29, App
30	122	4.4	845	10	US-09-738-626-4347	Sequence 4347, App
31	120	4.3	382	12	US-10-369-493-691	Sequence 691, App
32	120	4.3	382	12	US-10-369-493-15474	Sequence 15474, App
33	120	4.3	382	16	US-10-210-115-33	Sequence 33, App
34	116.5	4.2	656	12	US-10-369-493-18670	Sequence 18670, App
35	115.5	4.1	650	12	US-10-369-493-1338	Sequence 1338, App
36	115.5	4.1	650	12	US-10-369-493-20401	Sequence 20401, App
37	115.5	4.1	1687	14	US-10-094-679-3	Sequence 3, App
38	114	4.1	1071	12	US-10-369-493-23117	Sequence 23117, App
39	113.5	4.1	106	12	US-09-864-408A-2256	Sequence 2256, App
40	113	4.1	362	12	US-10-369-493-17332	Sequence 17332, App
41	112.5	4.0	371	12	US-10-369-493-11567	Sequence 11567, App
42	112.5	4.0	537	12	US-10-369-493-89	Sequence 89, App
43	111.5	4.0	728	12	US-10-369-493-15429	Sequence 15429, App
44	111.5	4.0	728	12	US-10-369-493-15797	Sequence 15797, App
45	111.5	4.0	728	12	US-10-369-493-16179	Sequence 16179, App

ALIGNMENTS

RESULT 1
US-09-934-903-12
Sequence 12, Application US/09934903
Patent No. US20020102690A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odem, J. Martin
APPLICANT: Schenzzel, Andreas J.
APPLICANT: No. US20020102690A1ton, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Roviére, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OR INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: C1646 US NA
CURRENT APPLICATION NUMBER: US/09/934, 903
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 12
LENGTH: 544
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF6
US-09-934-903-12

Query Match 100.0%; Score 2788; DB 10; Length 544;
Best Local Similarity 100.0%; Pred. No. 1.5e-254;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKFIFITGVVSSIGKXIAAIAIEDGKLVTKADPYINVDGTPSPFOHGVF 60
DB 1 MTKFIFITGVVSSIGKXIAAIAIEDGKLVTKADPYINVDGTPSPFOHGVF 60
QY 61 VTDEGATLDLIGHYERLTKTKKNNFTTGQVYEQVLRNRRGDYVATVQVPHITD 120

Db 61 VTGGAETDLDLGHYERLKTMTKNNFTTGQYVEQLANERKGDYIGATVQVPIHTD 120
QY 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFIHLTLVPYIK 180
Db 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFIHLTLVPYIK 180
QY 181 SAGELKTPTQSHVXELRTTIGIOPDILICRSEOPIPASERKIALFTVAERKAVISIDA 240
Db 181 SAGELKTPTQSHVXELRTTIGIOPDILICRSEOPIPASERKIALFTVAERKAVISIDA 240
QY 241 DTIYRIPLLREBQGLDLDVQDLRLDVPADLSAMERKVDGLTHPDEVSIAIVGKYVDH 300
Db 241 DTIYRIPLLREBQGLDLDVQDLRLDVPADLSAMERKVDGLTHPDEVSIAIVGKYVDH 300
QY 301 TDAYKSLNEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGGFEEGVEGK 360
Db 301 TDAYKSLNEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGGFEEGVEGK 360
QY 361 ISTVRFAENKIPYLGICLGMOSAVIEFARNVGLGASTETPLPSHPVIGLITEMD 420
Db 361 ISTVRFAENKIPYLGICLGMOSAVIEFARNVGLGASTETPLPSHPVIGLITEMD 420
QY 421 EAGELVTRDEDSDLDGTMRLGAQCKRLKADSLAFOLYQKDVITERRHRYEFNNQYIKOL 480
Db 421 EAGELVTRDEDSDLDGTMRLGAQCKRLKADSLAFOLYQKDVITERRHRYEFNNQYIKOL 480
QY 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPERTSTPRNGHALFSGFVEAAAKHT 540
Db 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPERTSTPRNGHALFSGFVEAAAKHT 540
QY 541 QGTA 544
Db 541 QGTA 544

RESULT 2

US-09-934-868-70
; Sequence 70, Application US/09934868
; Patent No. US2002017190A1
; GENERAL INFORMATION:
; APPLICANT: Odom, James M
; APPLICANT: Koffas, Mattheos
; APPLICANT: Odom, James M
; APPLICANT: Schenziel, Andreas J
; TITLE OF INVENTION: DENTRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 70
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; OTHER INFORMATION: Amino acid sequences encoded by ORF6 - PIRG
US-09-934-868-70

Query Match 100.0%; Score 2788; DB 10; Length 544;
Best Local Similarity 100.0%; Pred. No. 1.5e-254;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
Db 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
QY 61 VTGGAETDLDLGHYERLKTMTKNNFTTGQYVEQLANERKGDYIGATVQVPIHTD 120
Db 61 VTGGAETDLDLGHYERLKTMTKNNFTTGQYVEQLANERKGDYIGATVQVPIHTD 120
QY 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFIHLTLVPYIK 180

Db 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFIHLTLVPYIK 180
QY 181 SAGELKTPTQSHVXELRTTIGIOPDILICRSEOPIPASERKIALFTVAERKAVISIDA 240
Db 181 SAGELKTPTQSHVXELRTTIGIOPDILICRSEOPIPASERKIALFTVAERKAVISIDA 240
QY 241 DTIYRIPLLREBQGLDLDVQDLRLDVPADLSAMERKVDGLTHPDEVSIAIVGKYVDH 300
Db 241 DTIYRIPLLREBQGLDLDVQDLRLDVPADLSAMERKVDGLTHPDEVSIAIVGKYVDH 300
QY 301 TDAYKSLNEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGGFEEGVEGK 360
Db 301 TDAYKSLNEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGGFEEGVEGK 360
QY 361 ISTVRFAENKIPYLGICLGMOSAVIEFARNVGLGASTETPLPSHPVIGLITEMD 420
Db 361 ISTVRFAENKIPYLGICLGMOSAVIEFARNVGLGASTETPLPSHPVIGLITEMD 420
QY 421 EAGELVTRDEDSDLDGTMRLGAQCKRLKADSLAFOLYQKDVITERRHRYEFNNQYIKOL 480
Db 421 EAGELVTRDEDSDLDGTMRLGAQCKRLKADSLAFOLYQKDVITERRHRYEFNNQYIKOL 480
QY 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPERTSTPRNGHALFSGFVEAAAKHT 540
Db 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPERTSTPRNGHALFSGFVEAAAKHT 540
QY 541 QGTA 544
Db 541 QGTA 544

RESULT 3

US-09-941-947A-16
; Sequence 16, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Riviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Methylomonas 16a
US-09-941-947A-16

Query Match 100.0%; Score 2788; DB 11; Length 544;
Best Local Similarity 100.0%; Pred. No. 1.5e-254;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
Db 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
QY 61 VTGGAETDLDLGHYERLKTMTKNNFTTGQYVEQLANERKGDYIGATVQVPIHTD 120
Db 61 VTGGAETDLDLGHYERLKTMTKNNFTTGQYVEQLANERKGDYIGATVQVPIHTD 120

```

QY 121 EIKRKYVESAEKQVALLIEVGTVGDIIESLPLEETIRQMGVZLGRDRAFLTHLTVPIYK 180
Db 121 EIKRKYVESAEKQVALLIEVGTVGDIIESLPLEETIRQMGVZLGRDRAFLTHLTVPIYK 180
QY 181 SAGELEKTKPTQSHVSEKELRTIGIQPDLICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
Db 181 SAGELEKTKPTQSHVSEKELRTIGIQPDLICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
QY 241 DTIYIRIPILLREOGDLVDVLDLVPADLSAMEKVVDGLTHPTDEVSAIYVKYVDH 300
Db 241 DTIYIRIPILLREOGDLVDVLDLVPADLSAMEKVVDGLTHPTDEVSAIYVKYVDH 300
QY 301 TDVYSINBALIHAGIHRHKVOISYIDSETIEAGTALAKVVDAILVPGGGERGVGK 360
Db 301 TDVYSINBALIHAGIHRHKVOISYIDSETIEAGTALAKVVDAILVPGGGERGVGK 360
QY 361 ISTVFARENKIPYLGICLQMGSAVIEFARNVVGLEGASHTFLPKSPHPVIGLITEMWD 420
Db 361 ISTVFARENKIPYLGICLQMGSAVIEFARNVVGLEGASHTFLPKSPHPVIGLITEMWD 420
QY 421 ENGELVTRDESDLGSTRMLGAKCRKLKADSLAFQLYQKDVITERRHRRYEFNNQYLKOL 480
Db 421 ENGELVTRDESDLGSTRMLGAKCRKLKADSLAFQLYQKDVITERRHRRYEFNNQYLKOL 480
QY 481 EAAGKKSCKSLDGRLVIEILPEHPWFLACQFHPFTSTPRNGHALFSGFVEAAAKKH 540
Db 481 EAAGKKSCKSLDGRLVIEILPEHPWFLACQFHPFTSTPRNGHALFSGFVEAAAKKH 540
QY 541 OGTA 544
Db 541 OGTA 544

RESULT 4
US-09-815-242-11933
Sequence 11933, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Treawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO. 11933
LENGTH: 542
TYPE: PKT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11933

```

```

Query Match 72.0%, Score 2008, DB 9, Length 542;
Best Local Similarity 71.2%, Pred. No. 8,56-181;
Matches 383, Conservative 64, Mismatches 91, Indels 0, Gaps 0;

QY 1 MTKGIFITGGVSSLGKGIASSLAAILIEDGLKVTCTKDPYINVDGTMSPFOHGVF 60
Db 1 MTKGIFITGGVSSLGKGIASSLAAILIEDGLKVTCTKDPYINVDGTMSPFOHGVF 60
QY 61 VTEDGAEITDLGHYERFLTKTKNNFTTGQYVEOYLNERKQDYAGTAVIPIHTD 120
Db 61 VTEDGAEITDLGHYERFLTKTKNNFTTGQYVEOYLNERKQDYAGTAVIPIHTD 120
QY 121 EIKRKYVESAEKQVALLIEVGTVGDIIESLPLEETIRQMGVZLGRDRAFLTHLTVPIYK 180
Db 121 EIKRKYVESAEKQVALLIEVGTVGDIIESLPLEETIRQMGVZLGRDRAFLTHLTVPIYK 180
QY 181 SAGELEKTKPTQSHVSEKELRTIGIQPDLICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
Db 181 SAGELEKTKPTQSHVSEKELRTIGIQPDLICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
QY 241 DTIYIRIPILLREOGDLVDVLDLVPADLSAMEKVVDGLTHPTDEVSAIYVKYVDH 300
Db 241 DTIYIRIPILLREOGDLVDVLDLVPADLSAMEKVVDGLTHPTDEVSAIYVKYVDH 300
QY 301 TDVYSINBALIHAGIHRHKVOISYIDSETIEAGTALAKVVDAILVPGGGERGVGK 360
Db 301 TDVYSINBALIHAGIHRHKVOISYIDSETIEAGTALAKVVDAILVPGGGERGVGK 360
QY 361 ISTVFARENKIPYLGICLQMGSAVIEFARNVVGLEGASHTFLPKSPHPVIGLITEMWD 420
Db 361 ISTVFARENKIPYLGICLQMGSAVIEFARNVVGLEGASHTFLPKSPHPVIGLITEMWD 420
QY 421 ENGELVTRDESDLGSTRMLGAKCRKLKADSLAFQLYQKDVITERRHRRYEFNNQYLKOL 480
Db 421 ENGELVTRDESDLGSTRMLGAKCRKLKADSLAFQLYQKDVITERRHRRYEFNNQYLKOL 480
QY 481 EAAGKKSCKSLDGRLVIEILPEHPWFLACQFHPFTSTPRNGHALFSGFVEAAAKKH 538
Db 481 EAAGKKSCKSLDGRLVIEILPEHPWFLACQFHPFTSTPRNGHALFSGFVEAAAKKH 538
QY 538 EAAGKKSCKSLDGRLVIEILPEHPWFLACQFHPFTSTPRNGHALFSGFVEAAAKKH 538
Db 538 EAAGKKSCKSLDGRLVIEILPEHPWFLACQFHPFTSTPRNGHALFSGFVEAAAKKH 538

RESULT 5
US-09-815-242-10275
Sequence 10275, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Treawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16

```

NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for windows Version 4.0
SEQ ID NO 10275
LENGTH: 545
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10275

Query Match 68.7%; Score 1916; DB 9; Length 545;
Best Local Similarity 67.7%; Pred. No. 4.3e-172;
Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

```
QY 2 TKRIFITGVVSSLGKGIASAAIILDRGLKVTITKLDPIYINVDPGTMSPTQGEVTV 61
DB 3 TNYIFVTGVVSSLGKGIASAAIILDRGLKVTITKLDPIYINVDPGTMSPTQGEVTV 62
QY 62 TEGATETDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPHITDE 121
DB 63 TEGATETDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPHITNA 122
QY 122 IKRRVYESABEKDVALIEVGTVGDIESLPFLFTIRQKGVELGRDRALEPHILTVPIYKS 181
DB 123 IKRRVYESABEKDVALIEVGTVGDIESLPFLFTIRQKGVELGRDRALEPHILTVPIYMA 182
QY 182 AGELTKPTQHSVKELRTTIGIOPDILICRSEOPTPASBRKIALFTNVAEKAVISAIDAD 241
DB 183 SEEVKTRPTQHSVKELRTTIGIOPDILICRSDRAVPANERAKIALFCNVERKAVISLKDVD 242
QY 242 TIYRIPLLREOGDLDLVVDQLRLDVPAADSLAMEKVVYDGLTHPTDEVSIAIVKTYDHT 301
DB 243 SIYKIPGLKSGDLDVYICRFRFLCPPEANLSEWQVIFEEANPVSEVITIGWVGKYLEP 302
QY 302 DAKSLINALLHAGHTRHKVQISYIDSETIEASTALKNVDAILVVGSGERGVEGKI 361
DB 303 DAKSVINALLHAGHTRHKVQISYIDSETIEASTALKNVDAILVVGSGERGVEGKI 362
QY 362 STVRFARENKIPYLGICLQWQSAVIEFARNVVGLEGASTFELPKSPHPIVGLITEMODE 421
DB 363 TARFARENKIPYLGICLQWQSAVIEFARNVVGLEGASTFELPKSPHPIVGLITEMODE 422
QY 422 AGELVTRDESDLGTMRLGAKCKLADSLAPOLYXODVITERRHRYERFNNQYLKOLE 481
DB 423 NAEVTRDESDLGTMRLGAKCKLADSLAPOLYXODVITERRHRYERFNNQYLKOLE 482
QY 482 AAGKFSKSLDGRLEVEIIELEPHEWFLACQFHPFETSTPRNGHALFSGFVEAAK 537
DB 483 DAGLRVAGSGSDQVLVEIIELEPHEWFLACQFHPFETSTPRNGHALFSGFVEAAK 538
```

RESULT 6
US-09-815-242-11175
Sequence 11175, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for windows Version 4.0
SEQ ID NO 11175
LENGTH: 545
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-11175

Query Match 68.5%; Score 1908.5; DB 9; Length 545;
Best Local Similarity 67.5%; Pred. No. 2.2e-171;
Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

```
QY 2 TKRIFITGVVSSLGKGIASAAIILDRGLKVTITKLDPIYINVDPGTMSPTQGEVTV 61
DB 3 TNYIFVTGVVSSLGKGIASAAIILDRGLKVTITKLDPIYINVDPGTMSPTQGEVTV 62
QY 62 TEGATETDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPHITDE 121
DB 63 TEGATETDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPHITNE 122
QY 122 IKRRVYESABEKDVALIEVGTVGDIESLPFLFTIRQKGVELGRDRALEPHILTVPIYKS 181
DB 123 IKRRVYESABEKDVALIEVGTVGDIESLPFLFTIRQKGVELGRDRALEPHILTVPIYPT 182
QY 182 AGELTKPTQHSVKELRTTIGIOPDILICRSEOPTPASBRKIALFTNVAEKAVISAIDAD 241
DB 183 AGELTKPTQHSVKELRTTIGIOPDILICRSDRAVPANERAKIALFCNVERKAVISLKVYN 242
QY 242 TIYRIPLLREOGDLDLVVDQLRLDVPAADSLAMEKVVYDGLTHPTDEVSIAIVKTYDHT 301
DB 243 SIYKIPGLKSGDLDVYICRFRFLCPPEANLSEWQVIFEEANPVSEVITIGWVGKYLEP 302
QY 302 DAKSLINALLHAGHTRHKVQISYIDSETIEASTALKNVDAILVVGSGERGVEGKI 361
DB 303 DAKSVINALLHAGHTRHKVQISYIDSETIEASTALKNVDAILVVGSGERGVEGKI 362
QY 362 STVRFARENKIPYLGICLQWQSAVIEFARNVVGLEGASTFELPKSPHPIVGLITEMODE 421
DB 363 RTQVARENKIPYLGICLQWQSAVIEFARNVVGLEGASTFELPKSPHPIVGLITEMODE 422
QY 422 AGELVTRDESDLGTMRLGAKCKLADSLAPOLYXODVITERRHRYERFNNQYLKOLE 481
DB 423 EGVTEVTRDESDLGTMRLGAKCKLADSLAPOLYXODVITERRHRYERFNNQYLKOLE 482
QY 482 AAGKFSKSLDGRLEVEIIELEPHEWFLACQFHPFETSTPRNGHALFSGFVEAAK-AKHX 539
DB 483 KAGLVYGLSADKLEVEIIELEPHEWFLACQFHPFETSTPRNGHALFSGFVEAAK-AKHX 541
```

RESULT 7
US-09-815-242-13865
Sequence 13865, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A

```

/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 13865
/ LENGTH: 511
/ TYPE: PRT
/ ORGANISM: Salmonella typhi
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (1)...(511)
/ OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13865

```

```

Query Match      64.0%; Score 1783; DB 9; Length 511;
Best Local Similarity 67.4%; Pred. No. 1,5e-159;
Matches 341; Conservative 64; Mismatches 101; Indels 0; Gaps 0;

```

```

QY 2 TKRIFITGVSSLGKGIASSLAALIERGLKVTITKDPYINVDGTMSPFOHGEVY 61
DB 3 TNYIFVTVGVSSLGKGIASSLAALIERGLKVTITKDPYINVDGTMSPFOHGEVY 62
QY 62 TEDGATDLDLGHYERFLKTKTKKNNFTTGOYVEOVLNBRKGDYLGATVOVPHITD 121
DB 63 TEDGATDLDLGHYERFLKTKTKKNNFTTGOYVEOVLNBRKGDYLGATVOVPHITD 122
QY 122 IKRVRYSABKQVALLIEVGTVGDIESTPLETITQMGVEIGRPALEFIHTVPYIS 181
DB 123 IKRVRYSABKQVALLIEVGTVGDIESTPLETITQMGVEIGRPALEFIHTVPYIS 182
QY 182 AGELETKRPOHSHKELRTIGIOPDILICRSEOPIPASERKALFTNAEKAIVISAID 241
DB 183 AGELETKRPOHSHKELRTIGIOPDILICRSEOPIPASERKALFTNAEKAIVISAID 242
QY 242 TIYRIFPLLEGGDLVDLVVDLADVPADLSAMEKRVVDLTHPTDEVSAIVGKYVDHT 301
DB 243 SIYKIFGLKSGDLVDLVVDLADVPADLSAMEKRVVDLTHPTDEVSAIVGKYVDHT 302
QY 302 DAKSYLNEALIVGHTFRKVOISYIDSETIEAGTAKIKNDALIVPGGFGBRGVSGKI 361
DB 303 DAKSYLNEALIVGHTFRKVOISYIDSETIEAGTAKIKNDALIVPGGFGBRGVSGKI 362
QY 362 STVFRARENKIPYLGICLQKQSAVIEFPAENVGLBGAHSTIEFLPKSPHPVIGLIT 421
DB 363 STVFRARENKIPYLGICLQKQSAVIEFPAENVGLBGAHSTIEFLPKSPHPVIGLIT 422
QY 422 AGELVTRDESDLGIMRLGAOKCRLKASLSAQLYQKOVYTERHHRHREYFNQYIKOLE 481
DB 423 AGELVTRDESDLGIMRLGAOKCRLKASLSAQLYQKOVYTERHHRHREYFNQYIKOLE 482
QY 482 AAGMKFSGKSLDGRIVEIIELEPHEPW 507
DB 483 AAGMKFSGKSLDGRIVEIIELEPHEPW 508

```

RESULT 8
US-09-815-242-12372
; Sequence 12372, Application US/09815242
; Patent No. US20020061569A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: E117A.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 12372
/ LENGTH: 536
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-12372

```

```

Query Match      56.5%; Score 1576; DB 9; Length 536;
Best Local Similarity 55.6%; Pred. No. 6e-140;
Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

```

```

QY 1 MKRPIITGVSSLGKGIASSLAALIERGLKVTITKDPYINVDGTMSPFOHGEVY 60
DB 1 MKRPIITGVSSLGKGIASSLAALIERGLKVTITKDPYINVDGTMSPFOHGEVY 60
QY 61 VTDEGATDLDLGHYERFLKTKTKKNNFTTGOYVEOVLNBRKGDYLGATVOVPHITD 120
DB 61 VTDEGATDLDLGHYERFLKTKTKKNNFTTGOYVEOVLNBRKGDYLGATVOVPHITD 120
QY 121 EIKRVRYSABKQVALLIEVGTVGDIESTPLETITQMGVEIGRPALEFIHTVPYIS 178
DB 121 EIKRVRYSABKQVALLIEVGTVGDIESTPLETITQMGVEIGRPALEFIHTVPYIS 180
QY 178 IKSAGELTKRPOHSHKELRTIGIOPDILICRSEOPIPASERKALFTNAEKAIVISAID 238
DB 179 IKSAGELTKRPOHSHKELRTIGIOPDILICRSEOPIPASERKALFTNAEKAIVISAID 240
QY 238 DADITIRIPLLEGGDLVDLVVDLADVPADLSAMEKRVVDLTHPTDEVSAIVGKYVDHT 297
DB 239 DADITIRIPLLEGGDLVDLVVDLADVPADLSAMEKRVVDLTHPTDEVSAIVGKYVDHT 298
QY 297 DADITIRIPLLEGGDLVDLVVDLADVPADLSAMEKRVVDLTHPTDEVSAIVGKYVDHT 300
DB 298 DADITIRIPLLEGGDLVDLVVDLADVPADLSAMEKRVVDLTHPTDEVSAIVGKYVDHT 301
QY 300 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 360
DB 301 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 361
QY 360 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 420
DB 361 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 421
QY 420 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 476
DB 421 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 477
QY 476 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 536
DB 477 VSDIADIVSYVESLAKHAGYPPAKOVIDIRIMIDSSVYTBENNAEYLAADVIGIIVPGGF 536

```


Db 471 R0LEANGVAVISGTSPPDRLEWVEVIFPTNDPFIACQFHEPFLSRPWRPHIFKSFIEKSL 530
 QY 537 KHK 539
 Db 531 KYQ 533

RESULT 9

US-09-815-242-12978
 ; Sequence 12978, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant T.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12978
 ; LENGTH: 536
 ; TYPE: PRF
 ; ORGANISM: *Staphylococcus aureus*
 ; US-09-815-242-12978

Query Match 56.5%; Score 1576; DB 9; Length 536;

Best Local Similarity 55.6%; Pred. No. 6e-140; Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

QY 1 MTKRIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
 Db 1 MTKRIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
 QY 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGOVVEQVLRNKRKGDYLGATQVPIPHITD 120
 Db 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGOVVEQVLRNKRKGDYLGATQVPIPHITD 120
 QY 121 EIKRRVYSAB--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFLHLLTVPY 178
 Db 121 EIKRRVYSAB--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFLHLLTVPY 178
 QY 179 IKSAGELKPTQHSVKELERTIGIOPDILICRSEOPFASERKRIALFTVNAEKAVISAI 238
 Db 181 IKSAGELKPTQHSVKELERTIGIOPDILICRSEOPFASERKRIALFTVNAEKAVISAI 238
 QY 239 DADITVYPIPLKEQGLDLDVQDLRDV--AADLSMEKVDGLTHFTDVSIAIVGK 297
 Db 241 DADISLYEIPQLSQQNDVDIVYIKRLQINAKYETQLDDEWKQLDIYNNLDGKITITGLVGK 300
 QY 298 VDHTEADKSLNEALIHGHIHTRHKVQISYIDSEITIEAGTAK--LKNVDAILVPGGREG 356

Db 301 VSLQDAKISVSESILKHGYPPAKDIDIRWIDSSSVTDEMAEVLADVDGILVPGGFRRA 360
 QY 357 VEKGISFVRFARENKIFPIGLICLQMSQAVIEFANVVGLEGASHTEELPNSPHFVIGLIT 416
 Db 361 SEGKISAKTVARENNVPPFGICLQMLATVFPNSNVGLBGAHSELDPAFPYPIIDLP 420
 QY 417 EMDPAGELVTRDDSDSLGTRLGAKCRKADSLAFOLYQKVITERHRRHREFFNNQY 476
 Db 421 EQKD-----IEDLGGTRLGLYPCSIXEGTLAQDVYGAIEERHRRHREFFNNQY 470
 QY 477 LKOLBAGMKFSGKSLQGRVLEIIELEPHWPFLACQFHEPFTSPRGHALFSGFVAAA 536
 Db 471 R0LEANGVAVISGTSPPDRLEWVEVIFPTNDPFIACQFHEPFLSRPWRPHIFKSFIEKSL 530
 QY 537 KHK 539
 Db 531 KYQ 533

RESULT 10

US-09-815-242-4984
 ; Sequence 4984, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant T.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4984
 ; LENGTH: 536
 ; TYPE: PRF
 ; ORGANISM: *Enterococcus faecalis*
 ; US-09-815-242-4984

Query Match 56.3%; Score 1569; DB 9; Length 536;

Best Local Similarity 55.4%; Pred. No. 2.8e-139; Matches 301; Conservative 94; Mismatches 134; Indels 14; Gaps 4;

QY 1 MTKRIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
 Db 1 MTKRIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSPFQHGVEF 60
 QY 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGOVVEQVLRNKRKGDYLGATQVPIPHITD 120
 Db 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGOVVEQVLRNKRKGDYLGATQVPIPHITD 120
 QY 121 EIKRRVYSAB--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFLHLLTVPY 178
 Db 121 EIKRRVYSAB--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFLHLLTVPY 178

Db	121	ETKEKIMPAKKTADAVITTEVGVGDIESLPLEBALRQKKAQMGSDNWMYHTITLIFY	180
Oy	179	IKSAGELTKKPTQHSVKELRTIGIOPDILICRSEQPIASERRKIALFTVAEKAVISAL	238
Db	181	LKQKGEKTKPTQHSVKELRSIGIQPNILVTRTELPSQNTKMLAQFCOVNPEAVI ESR	240
Oy	239	DADTVIRIFELLREQGLDLDVVDQJLZDVPAADLSAMERKVVYDGLTHPTDEVSIAIVKTY	298
Db	241	DVEITLVSIFLALQNMQDQICDHKLADAPADWTENRALREKYNLKKKTKYALVERKY	300
Oy	299	DHTDAKSLNEALIHAGITREKQVSIYSDSTIABGTA-KLKNVDAIIVPGSGERG	357
Db	301	ELPDAYISVEVALKHAGFDPDSDEIELDWDSQELTAENVARISAGIILVPGSGRG	360
Oy	358	BKGISVTRFARENKPYLGICIGMOSAVIEFARNVGLGASHTEFLPKSPHPYIGLITE	417
Db	361	BKITEAIRARENDVFPJGICIGMOMACYEERBNVVGLEDQSAFTNDVYNNIIDLMAD	420
Oy	418	WMDEAGELVTRDESDJLGGTMYLGAQCKMLKADSLAFOLY-QQVYTERHRHRYENNQY	476
Db	421	-----QENIEMLGGLTRIGLYPCRLKKGTATAAYANGMEDVQRRHHRHREYFNKKY	4707
Oy	477	LKOLEAAGKFKSKSLDGRVLVETIELPEHPWFLAQERPEFTSPRNCHALFSGFVAAA	536
Db	471	ROLFEENGIVFSGVSPDNRLVETIVEIPEKQFVAAQZPELISRNPRQRLIKQFVGAAL	530
Oy	537	KHK 539	
Db	531	ANK 533	

```

RESULT 11
US-09-815-242-10727
? Sequence 10727, Application US/09815242
? Patent No. US2002061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Kari L.
? APPLICANT: Zykkind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant U.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes in
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: EILITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/1191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: fastseq for windows Version 4.0
? SEQ ID NO 10727
? LENGTH: 536
? TYPE: PRT
? ORGANISM: Enterococcus faecalis
US-09-815-242-10727
Query Match
56.3%, Score 1569, DB 9, Length 536

```

[illegible]

```

1 RESULT 12
2
3 ; Sequence 13376, Application US/09815242
4 ; Patent No. US20020061569A1
5
6 GENERAL INFORMATION:
7
8 APPLICANT: Haselbeck, Robert
9
10 APPLICANT: Ohlsem, Kari L.
11
12 APPLICANT: Zvekind, Judith W.
13
14 APPLICANT: Wall, Daniel
15
16 APPLICANT: Trewick, John D.
17
18 APPLICANT: Carr, Grant J.
19
20 APPLICANT: Yamamoto, Robert T.
21
22 APPLICANT: Xu, H. Howard
23
24 TITLE OF INVENTION: Identification of Essential Genes in
25 TITLE OR INVENTION: Prokaryotes
26
27 FILE REFERENCE: ELITRA.0114
28
29 CURRENT APPLICATION NUMBER: US/09/815,242
30
31 CURRENT FILING DATE: 2001-03-21
32
33 PRIOR APPLICATION NUMBER: 60/119,078
34
35 PRIOR FILING DATE: 2000-03-21
36
37 PRIOR APPLICATION NUMBER: 60/206,848
38
39 PRIOR FILING DATE: 2000-05-23
40
41 PRIOR APPLICATION NUMBER: 60/207,727
42
43 PRIOR FILING DATE: 2000-05-26
44
45 PRIOR APPLICATION NUMBER: 60/242,578
46
47 PRIOR FILING DATE: 2000-10-23
48
49 PRIOR APPLICATION NUMBER: 60/253,625
50
51 PRIOR FILING DATE: 2000-11-27
52
53 PRIOR APPLICATION NUMBER: 60/257,931

```

PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 13376
 LENGTH: 535
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-815-242-13376

Query Match 56.0%; Score 1561; DB 9; Length 535;
 Best Local Similarity 54.3%; Pred. No. 1,66-138;
 Matches 294; Conservative 108; Mismatches 125; Indels 14; Gaps 4;

2 TKFPIITGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGFEVY 61
 3 TKYIFVTGGVSSIGKGIASSLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGFEVY 62
 62 TEDGATDLDGHERFFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDE 121
 63 TDDGATDLDGHERFFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDA 122
 122 IKRRYESA--EGKVALTEVGTVGDIESTLPLETIRQMGVELGRBALFHLTVPEYI 179
 123 LKFKIKRALTTDSVITTEVGTVGDIESTLPLETIRQMGVELGRBALFHLTVPEYL 182
 180 KSAGELKTPQHSYKELRTIGIOPDILCSSEQIPASERKILFTNVAEKAVISAD 239
 183 KKAAGMKKPTQHSYKELRTIGIOPDILCSSEQIPASERKILFTNVAEKAVISAD 242
 240 ADTIRIPLLRBQGLDLDVVDQLRDLVPAADLSAMEKVDGLTPTDEVSIAIVGYVD 299
 243 VEHVQIPLNLAQGMDOYVCHLKDAPADMTMSAMVDKVMKLVKQVYISLVKGYE 302
 300 HTDAVKSLEALIHAGIHTRHVKQISYIDSETTEAEGLNK--LKNVDAILVPGFGEKGYE 358
 303 LQDAVYSVEALKHSGYVNDVEVKINWVANNDVTAEVAILSDADGIVPGFGQORSTE 362
 359 GKTSTWRFRKPKIPYIGICLGMQSAVIEFANVVGLEGAHSTFELPSHPHVLITDEM 418
 363 GKIQIRKRENDVPRLGVCIDMOLCTIEFAHVGLBGSANBALAPETKVIYIIMDQ 422
 419 MDAGELVTRDESDLGGMRLGAQKCRILKADSLAFQLY--QKVITERRHRYEENQYL 477
 423 ID-----IDMGCTLRIGLVPSKLRGSKKAAAHVNOEYVQRHRRHYEENNAFR 472
 478 KQLEAGMTPSGKSLDGRVLEIIEPEHPWFLACQHPFETSTPNNGHALFSGFEAAAX 537
 473 EGFPAAGFVFSVPNRLVLEIIEPEHPWFLACQHPFETSTPNNGHALFSGFEAAAX 532
 QY 538 H 538
 DB 533 N 533

RESULT 13
 US-09-815-242-11320
 Sequence 11320, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kati L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 11320
 LENGTH: 538
 TYPE: PRT
 ORGANISM: Helicobacter pylori
 US-09-815-242-11320

Query Match 51.8%; Score 1444.5; DB 9; Length 538;
 Best Local Similarity 50.8%; Pred. No. 1,76-127;
 Matches 271; Conservative 112; Mismatches 147; Indels 3; Gaps 2;

3 KPIFTGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGFEVY 62
 5 KPIFTGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGFEVY 64
 63 EDGATDLDGHERFFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDEI 122
 65 SDGATDLDGHERFFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDEI 124
 123 KRRYESAEGVDVALLTEVGTVGDIESTLPLETIRQMGVELGRBALFHLTVPEYKSA 182
 125 KRRKSAKGIPLTEVGTVGDIESTLPLETIRQMGVELGRBALFHLTVPEYKSA 184
 183 GELTKPTQHSYKELRTIGIOPDILCSSEQIPASERKILFTNVAEKAVISADDT 242
 185 NELTKPTQHSYKELRTIGIOPDILCSSEQIPASERKILFTNVAEKAVISADDT 244
 243 IYRPLLRBQGLDLDVVDQLRDLVPAADLSAMEKVDGLTPTDEVSIAIVGYVDHT 302
 245 IYAPPIILFQGLITPARRFNKLPKMAAMNTLVKTIAPKAKYIGVGYLSIKE 304
 303 AYKSLNEALIHAGIHTRHVKQISYIDSETTEAEGLNK--LKNVDAILVPGFGEKGYE 362
 305 SYKSLIEALIHAGIHTRHVKQISYIDSETTEAEGLNK--LKNVDAILVPGFGEKGYE 362
 363 TWRPAREKTIPLYIGICLGMQSAVIEFANVVGLEGAHSTFELPSHPHVLITDEMDEA 422
 363 AIQARLEKLPYIGICLGMQSAVIEFANVVGLEGAHSTFELPSHPHVLITDEMDEA 422
 423 GELVTRDESDLGGMRLGAQKCRILKADSLAFQLYQKQVITERRHRYEENQYLKQLEA 482
 423 HQQVRTYNSPLGTMPLGRCIMNSLLEKAYKPSIKERHRYEENQYLKQLEA 482
 483 AGMTPSGKSLDGRVLEIIEPEHPWFLACQHPFETSTPNNGHALFSGFEAAAX 535
 DB 483 KGLKVVGG--SNHLIEALIEHDHPFVGVQFHPETSTPNNGHALFSGFEAAAX 534

RESULT 14
 US-10-156-761-14031
 Sequence 14031, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
 APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHITAKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 14031
 LENGTH: 549
 TYPE: PRT
 ORGANISM: Streptomyces avermitilis
 US-10-156-761-14031

Query Match 51.4%; Score 1432.5; DB 15; Length 549;
 Best Local Similarity 52.5%; Pred. No. 2.3e-126;
 Matches 287; Conservative 92; Mismatches 153; Indels 15; Gaps 7;

QY 2 TKRIFITGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 61
 DB 8 TKHIFVTGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 67
 QY 62 TEDGATDLDIGHERFLKTKMTKKNFTTGOVVEOVLANERKGDYLGATVQVIPHIDE 121
 DB 68 TNDGATDLDIGHERFLKTKMTKKNFTTGOVVEOVLANERKGDYLGATVQVIPHIDE 127
 QY 122 IKRRVYESAEG-DVALIEVGTVGDIESLPLETTIRONGVEIGRDLALFIHLTVPIYK 180
 DB 128 IKRRIRKATDEVDVITVEGVTVGDIESLPLETTIRONGVEIGRDLALFIHLTVPIYK 187
 QY 181 SAGELKTKPTQSVKELRTIGTQDILIRSEOPIPASERKIALFTNVAEKAVISAIDA 240
 DB 188 PEGELKTKPTQSVKELRTIGTQDILIRSEOPIPASERKIALFTNVAEKAVISAIDA 247
 QY 241 DTYIRPLLEIRGQDLDLVVDOLRLDVPADLSAMEKVVDTLPTDEVSIAIVGYVHD 300
 DB 248 RSTIYDFKTVHGGDLDVYVVKLDLPFRVDWITWMDLLDRHKPDHEINLAVGKIDL 307
 QY 301 TDAYSLSNALIHAGIHTHRKQVQISYIDSEITL-ABGTAKLKNVDAILVPGFBERGYE 358
 DB 308 PAVYLSVTALRAAGFANRPAVYIKWTSDDCKTPAGAGQGLGDVDAIICPGFGRGVS 367
 QY 359 GKISTVPAREKTIPTLIGICIGQSAVIEFARNVGLLEGASTETPLPKSPHPIGLITEM 418
 DB 368 GKVGAIQVAREKTIPTLIGICIGQSAVIEFARNVGLLEGASTETPLPKSPHPIGLITEM 427
 QY 419 MD-EAGELVTRDESDLGTMRLGAQCKRLKADSLAPOLYO-KDVTERRHRYEPNNQY 476
 DB 428 LDIIVAGE-----GDMGTMRLGWPAKLAECSIVREYDGEYVEERHRYEYNNAY 480
 QY 477 LKOLE-AAGCKFSKSLDGRIVEIIELEP--HPWFLACQFPBEPSTPRNGALSSGPVE 533
 DB 481 RALEKKAQGLQSSGTSBDGLVEYVAPREVPYVATQAHPELSSRPTRPPLPAGLVK 540
 QY 534 AAKHKT 540
 DB 541 AAVERTK 547

RESULT 15
 US-09-738-626-5066
 Sequence 5066, Application US/09738626
 Publication No. US20020197605A1
 GENERAL INFORMATION:
 APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO
 APPLICANT: YATEISHI, NAOKO
 APPLICANT: SENO, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 CURRENT FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SEQ ID NO 5066
 SOFTWARE: Patent in ver. 3.0
 LENGTH: 554
 TYPE: PRT
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-5066

Query Match 49.2%; Score 1372; DB 10; Length 554;
 Best Local Similarity 51.9%; Pred. No. 1.3e-120;
 Matches 283; Conservative 78; Mismatches 166; Indels 18; Gaps 6;

QY 2 TKRIFITGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 61
 DB 10 TKHIFVTGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 69
 QY 62 TEDGATDLDIGHERFLKTKMTKKNFTTGOVVEOVLANERKGDYLGATVQVIPHIDE 121
 DB 70 TEDGATDLDIGHERFLKTKMTKKNFTTGOVVEOVLANERKGDYLGATVQVIPHIDE 129
 QY 122 IKRRVYESAEG-DVALIEVGTVGDIESLPLETTIRONGVEIGRDLALFIHLTVPIYK 175
 DB 130 IKRRIRKATDEVDVITVEGVTVGDIESLPLETTIRONGVEIGRDLALFIHLTVPIYK 189
 QY 176 VPIYKSAEELKTKPTQSVKELRTIGTQDILIRSEOPIPASERKIALFTNVAEKAVISAIDA 235
 DB 190 VPIYKSAEELKTKPTQSVKELRTIGTQDILIRSEOPIPASERKIALFTNVAEKAVISAIDA 249
 QY 236 SAIDADTYIRPLLEIRGQDLDLVVDOLRLDVPADLSAMEKVVDTLPTDEVSIAIVGYVHD 295
 DB 250 SCPOSSISYINIPDYLREHLDPTFIIRLGLPFRVDWITWMDLLDRHKPDHEINLAVGKIDL 309
 QY 296 KYVDHTDAYSLSNALIHAGIHTHRKQVQISYIDSEITL-ABGTAKLKNVDAILVPGFBERGYE 353
 DB 310 KYVDHTDAYSLSNALIHAGIHTHRKQVQISYIDSEITL-ABGTAKLKNVDAILVPGFBERGYE 369
 QY 354 ERGVGKISTVPAREKTIPTLIGICIGQSAVIEFARNVGLLEGASTETPLPKSPHPIGLITEM 413
 DB 370 IRGIBGKIGATFPAREKTIPTLIGICIGQSAVIEFARNVGLLEGASTETPLPKSPHPIGLITEM 428
 QY 414 LITWMDAAGELVTRDESDLGTMRLGAQCKRLKADSLAPOLYO-KDVTERRHRYEPNNQY 473
 DB 429 TMBE-----OKAAVSGADLDGTMRLGAQCKRLKADSLAPOLYO-KDVTERRHRYEPNNQY 482
 QY 474 NOYLKOL-EAAGCKFSKSLDGRIVEIIELEP--HPWFLACQFPBEPSTPRNGALSSGPVE 530
 DB 483 NAYRAQIIEGSDLVFSGTSPGHVVEYVAPREVPYVATQAHPELSSRPTRPPLPAGLVK 542
 QY 531 FVEAA 535
 DB 543 LVKTA 547

Search completed: January 29, 2004, 16:21:18
 Job time : 38.1346 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 16.0063 Seconds
(without alignments)
3268.453 Million cell updates/sec

Title: US-09-941-947a-16

Perfect score: 2788

Sequence: 1 MTKRFFITGGVSSLGKGLA.....HALPSGFVEAAAKHTQCTA 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	72.0	542	B83192	CTP synthase PA363
2	1946	69.8	545	AD0859	CTP synthetase [im
3	1894	69.0	545	E82074	CTP synthase VC244
4	1916	68.7	545	1	CTP synthetase (EC 6
5	1916	68.7	545	1	CTP synthetase (im
6	1916	68.7	545	1	CTP synthetase (im
7	1909	68.5	545	2	CTP synthetase (EC 6
8	1908.5	68.5	545	2	CTP synthetase (EC 6
9	1905	68.3	554	2	CTP synthetase XPI
10	1892.5	67.9	544	2	CTP synthetase (EC 6
11	1888.5	67.7	544	2	CTP synthetase (EC 6
12	1737	62.3	545	2	CTP synthetase (EC 6
13	1645.5	59.0	532	2	CTP synthetase (EC 6
14	1639	58.8	531	2	CTP synthetase - A
15	1622.5	57.9	535	1	CTP synthetase (EC 6
16	1615.5	57.9	535	1	CTP synthetase (UTP-
17	1603.5	57.5	552	2	CTP synthetase (EC 6
18	1600.5	57.4	532	2	CTP synthetase (EC 6
19	1600.5	57.4	532	2	CTP synthetase (EC 6
20	1595	57.2	550	2	CTP synthetase (EC 6
21	1592	57.1	542	2	CTP synthetase (EC 6
22	1576	56.5	536	2	CTP synthetase (EC 6
23	1574.5	56.5	545	2	CTP synthetase (impo
24	1563.5	56.1	540	2	CTP synthetase (impo
25	1562	56.0	535	2	CTP synthetase (EC 6
26	1562	56.0	535	2	CTP synthetase (impo
27	1562	56.0	535	2	CTP synthetase (impo
28	1561	56.0	535	2	CTP synthetase (UTP-
29	1555.5	55.8	537	2	CTP synthetase (pyrg

30	1542	55.3	544	2	825101	CTP synthase (EC 6
31	1538.5	55.2	557	2	H71189	probable CTP synth
32	1507	54.1	543	2	E81418	CTP synthetase (EC 6
33	1484	53.2	533	2	E69154	CTP synthetase - Met
34	1449	52.0	538	2	C71946	CTP synthetase - Hel
35	1444.5	51.8	538	2	E64563	CTP synthetase - H
36	1441.5	51.7	549	2	T36879	probable CTP synth
37	1434	51.4	535	2	C86686	CTP synthetase (im
38	1423	51.0	544	2	B75379	CTP synthetase - Del
39	1374.5	49.3	532	2	D69281	CTP synthetase (pyrg
40	1369.5	49.1	577	2	C71342	probable CTP synth
41	1365.5	49.0	586	2	B70503	probable pyrg prot
42	1362	48.9	553	2	B84334	CTP synthetase (impo
43	1356.5	48.7	537	2	A72103	CTP synthetase - C
44	1356	48.6	535	2	D90161	CTP synthetase (UT
45	1355.5	48.6	537	2	D86520	CTP synthetase (im

ALIGNMENTS

RESULT 1

B83192 CTP synthase PA3637 [imported] - Pseudomonas aeruginosa (strain PA01)

C.Species: Pseudomonas aeruginosa

C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C.Accession: B83192

R.Stover, C.K., Pham, X.O., Erwin, A.L., Mizoguchi, S.D., Warener, P., Hickey, M.J., Bri

adman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Laidig, K., Lim,

, Loir, S., Olson, M.V.

Nature 406, 959-964, 2000

A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoc

A.Reference number: A82950; MUID:20437337; PMID:10984043

A.Accession: B83192

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-542 <STD>

A.Cross-references: GB:AE004783; GB:AE004091; MID:9949786; PIDN:AG07025.1; GSPDB:GNO01:

A.Experimental source: strain PA01

C.Genetics:

A.Gene: pyrg: PA3637

C.Superfamily: CTP synthase

Query Match 72.0%; Score 2008; DB 2; Length 542;

Best Local Similarity 71.2%; Pred. No. 3.1e-121;

Matches 383; Conservative 64; Mismatches 91; Indels 0; Gaps 0;

QY	1	MTKRFFITGGVSSLGKGLAASSIAATLEDRGLKVTITKIDPYINVDPTWSPFGHGVF	60
DB	1	MTKRFFITGGVSSLGKGLASASIAATLEDRGLKVTITKIDPYINVDPTWSPFGHGVF	60
QY	61	VTBGAFTDLDLGHYERFLKTTMTKKNFTTGQVYEVOLNREKGDYLGATVOYIPHTD	120
DB	61	VTBGAFTDLDLGHYERFLKTTMTKKNFTTGQVYEVOLNREKGDYLGATVOYIPHTD	120
QY	121	EIKRRVYSAGKQVALIEVGCTGDTESLPFTTIOMGYELDRDALFHTLTPPIK	180
DB	121	EIKRRVYSAGKQVALIEVGCTGDTESLPFTTIOMGYELDRDALFHTLTPPIK	180
QY	181	SAGELKTPQHSVKEKRTIGIOPDILICRSEOPTPASERKIALFTNVAEKAVISAIDA	240
DB	181	SAGELKTPQHSVKEKRTIGIOPDILICRSEOPTPASERKIALFTNVAEKAVISAIDA	240
QY	241	DTIIRPLLEEGGLDIDVDQLRLDVPADLSAMEKYVDGLHPTDVSIAIYKIVDH	300
DB	241	DTIIRPLLEEGGLDIDVDQLRLDVPADLSAMEKYVDGLHPTDVSIAIYKIVDH	300
QY	301	TDVAKSLNEALIHGITHKRVQSYIDSEITIEAGTAKLVNDALIVPGSGRGVEGK	360
DB	301	TDVAKSLNEALIHGITHKRVQSYIDSEITIEAGTAKLVNDALIVPGSGRGVEGK	360
QY	361	ISTVFAREKRPYLGICLQKQSAVIFPARNVGLGASHTEFLPKSPHYIGIITMD	420
DB	361	ISTVFAREKRPYLGICLQKQSAVIFPARNVGLGASHTEFLPKSPHYIGIITMD	420

Db 361 ISTVQYARBNKIPYIGICLGMQVAVIYARNVIGSDANSTFDSKSGHPVALLTEWMD 420

QY 421 EAGELVTRDESDLDGGTMRGAOKCRLKADSLAFOLYQKOVITERRRHRYEFNNQYLKOL 480

Db 421 AAGATGTRFTRADSLGGTMRGAQEQCLQGTCLVHOCYADVAVENRRHRYEFNNMLLPOL 480

QY 481 EAAAGKFSKSLDGRVLEIIELEPHFWFLACQHPHFTSTPRNGALFSGFVBAAKH 538

Db 481 EAGAKTISGRSDGALVEVFAPEHFWFVACQHPHFTSTPRDGHPLSGFVBAALMKY 538

RESULT 2

AD0859

CTP synthetase (imported) - Salmonella enterica subsp. enterica serovar Typhi (strain CT C/Specties: Salmonella enterica subsp. enterica serovar Typhi

A>Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AD0859

R:Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Commerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Koul, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AD0859

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-545 <PAR>

A/Cross-References: GB:AL513382; PIDN:CAD06059.1; PID:g16504026; GSPDB:GN00176

C/Genetics:

A:Gene: STY3082

C/Superfamily: CTP synthase

Query Match 69.8%; Score 1946; DB 2; Length 545;
Best Local Similarity 68.2%; Pred. No. 36-117;
Matches 367; Conservative 73; Mismatches 98; Indels 0; Gaps 0;

QY 2 TKFIFITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDPGTMSPGHGVFV 61

Db 3 TNYIFITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDPGTMSPGHGVFV 62

QY 62 TEDGATDLDGHERFLKTYTKKNNFTGGVYEQVLENERKDYLGATVQVIPHITDE 121

Db 63 TEDGATDLDGHERFLKTYTKKNNFTGGVYEQVLENERKDYLGATVQVIPHITNA 122

QY 122 IKRKYVESAEKQVALLIEVGTVGDIESTLPLEETIRQMGVELGRDALFIHLTLVPIYKS 181

Db 123 IKRKYVESAEKQVALLIEVGTVGDIESTLPLEETIRQMGVELGRDALFIHLTLVPIYAA 182

QY 182 AGEVKTPTQHSVKEKLTIGIOPDILICRSEQIPASERKIALFTVAERKAVISAIDAD 241

Db 183 AGEVKTPTQHSVKEKLTIGIOPDILICRSEQIPASERKIALFTVAERKAVISMKDVD 242

QY 242 TYIRPILLREOGDLDLVQDLRDVPAADLSAMKRVVDGLTHPTDEVSIAVGYVDHT 301

Db 243 STYKIPOLISQGLDLYICRPSLNCPEKNSMQLVYEEZANPAGSVITIGVGYTILP 302

QY 302 DAYKSLNEALIHAGIHTRHVKQISYIDSETTEAGETAKLNVDAILVPGGGERGVESKI 361

Db 303 DAYKSLNEALIHAGIHTRHVKQISYIDSETTEAGETAKLNVDAILVPGGGERGVESKI 362

QY 362 STVPRARENKIPYIGICLGMQSAVIEFARNVVGLEGASTFLPSHPVIGLITEMWDE 421

Db 363 STVPRARENKIPYIGICLGMQSAVIEFARNVVGLEGASTFLPSHPVIGLITEMWDE 422

QY 422 AGEVLTREDSDLDGGTMRGAOKCRLKADSLAFOLYQKOVITERRRHRYEFNNQYLKOL 481

Db 423 DQAVETRSKSLDGGTMRGAQEQCLQSDSLVRLQYAGSTIYERRHRYEFNNMLLPOL 482

QY 482 AAGKFSKSLDGRVLEIIELEPHFWFLACQHPHFTSTPRNGALFSGFVBAAKH 539

Db 483 AAGKFSKSLDGRVLEIIELEPHFWFLACQHPHFTSTPRNGALFSGFVBAAKH 540

RESULT 3

E82074

CTP synthase VC2448 (imported) - Vibrio cholerae (strain N16961 serogroup O1)

C/Specties: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C/Accession: E82074

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.

1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406 477-483, 2000

A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: AB0395; MUID:20406833; PMID:10952301

A/Accession: E82074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-545 <HRI>

A/Cross-References: GB:AE004315; GB:AE003852; MUID:9657024; PIDN:AA955590.1; GSPDB:GN001

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

C/Genetics:

A:Gene: VC2448

A:Map position: 1

C/Superfamily: CTP synthase

Query Match 69.0%; Score 1924; DB 2; Length 545;
Best Local Similarity 67.5%; Pred. No. 7-76-116;
Matches 362; Conservative 76; Mismatches 98; Indels 0; Gaps 0;

QY 2 TKFIFITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDPGTMSPGHGVFV 61

Db 3 TNYIFITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDPGTMSPGHGVFV 62

QY 62 TEDGATDLDGHERFLKTYTKKNNFTGGVYEQVLENERKDYLGATVQVIPHITDE 121

Db 63 TEDGATDLDGHERFLKTYTKKNNFTGGVYEQVLENERKDYLGATVQVIPHITNA 122

QY 122 IKRKYVESAEKQVALLIEVGTVGDIESTLPLEETIRQMGVELGRDALFIHLTLVPIYKS 181

Db 123 IKRKYVESAEKQVALLIEVGTVGDIESTLPLEETIRQMGVELGRDALFIHLTLVPIYAA 182

QY 182 AGEVKTPTQHSVKEKLTIGIOPDILICRSEQIPASERKIALFTVAERKAVISAIDAD 241

Db 183 AGEVKTPTQHSVKEKLTIGIOPDILICRSEQIPASERKIALFTVAERKAVISMKDVD 242

QY 242 TYIRPILLREOGDLDLVQDLRDVPAADLSAMKRVVDGLTHPTDEVSIAVGYVDHT 301

Db 243 STYKIPOLISQGLDLYICRPSLNCPEKNSMQLVYEEZANPAGSVITIGVGYTILP 302

QY 302 DAYKSLNEALIHAGIHTRHVKQISYIDSETTEAGETAKLNVDAILVPGGGERGVESKI 361

Db 303 DAYKSLNEALIHAGIHTRHVKQISYIDSETTEAGETAKLNVDAILVPGGGERGVESKI 362

QY 362 STVPRARENKIPYIGICLGMQSAVIEFARNVVGLEGASTFLPSHPVIGLITEMWDE 421

Db 363 STVPRARENKIPYIGICLGMQSAVIEFARNVVGLEGASTFLPSHPVIGLITEMWDE 422

QY 422 AGEVLTREDSDLDGGTMRGAOKCRLKADSLAFOLYQKOVITERRRHRYEFNNQYLKOL 481

Db 423 EGVNEESSEKSLDGGTMRGAQEQCLQGTCLVHOCYADVAVENRRHRYEFNNMLLPOL 482

QY 482 AAGKFSKSLDGRVLEIIELEPHFWFLACQHPHFTSTPRNGALFSGFVBAAKH 537

Db 483 AAGKFSKSLDGRVLEIIELEPHFWFLACQHPHFTSTPRNGALFSGFVBAAKH 538

RESULT 4

SECTP

CTP synthase (EC 6.3.4.2) (validated) - Escherichia coli (strain K-12)

N/Alternate names: CTP-synthetase; UTP-ammonia ligase

C/Specties: Escherichia coli

C/Date: 31-Mar-1998 #sequence_revision 21-Nov-1997 #text_change 01-Mar-2002

C/Accession: H65059; A25608

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A>Title: The complete genome sequence of *Escherichia coli* K-12.
 A;Reference number: A64720; MUID:9742617; PMID:9278503
 A;Accession: H65059
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A;Residues: 1-545 <BLAT>
 A;Cross-references: GB:AE000361; GB:U00096; NID:92367160; PIDN:AACT5822.1; PID:91789142
 A;Experimental source: strain K-12, substrain MG1655
 R;Meng, M.; Makarov, C.A.; Zalkin, H.
 J. Biol. Chem. 261, 5568-5574, 1986
 A>Title: Nucleotide sequence of *Escherichia coli* pyrg encoding CTP synthetase.
 A;Reference number: A92584; MUID:86168304; PMID:3514618
 A;Accession: A25608
 A:Molecule type: DNA
 A;Residues: 1337,'L',339-475,'S',477-488,'RA',491-545 <MEN>
 A;Cross-references: GB:M12843
 C;Genetics:
 A;Gene: pyrg
 A;Map position: 60 min
 C;Function: homodimer; aggregates to a tetramer
 A;Description: this glutamine amidotransferase catalyzes the terminal reaction in the de
 C;Superfamily: CTP synthase
 C;Keywords: homodimer; 1-gase; pyrimidine nucleotide biosynthesis
 F;2-545/Product: CTP synthase #status experimental <MAT>

Query Match 68.7%; Score 1916; DB 1; Length 545;
 Best Local Similarity 67.7%; Pred. No.2.5e-115;

Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

QY 2 TKEFIITGGVSSLSKGIAGASLSAILEDGKVTITKLDPIYNDPGTMSPOHGEVY 61
 DB 3 TTYIFVTGGVSSLSKGIAGASLSAILEDGKVTITKLDPIYNDPGTMSPOHGEVY 62
 QY 62 TEDGATDLDGHTERFRTKMTKKNFTTGOVYEQVLENERKDYLGATVQVPHITDE 121
 DB 63 TEDGATDLDGHTERFRTKMTKKNFTTGOVYEQVLENERKDYLGATVQVPHITDE 122
 QY 122 IKRRVYASAEKQVALIEVGTGDIESTPLETIRONGVLEGRDALPHILTLVPYKIS 181
 DB 123 IKRRVYASAEKQVALIEVGTGDIESTPLETIRONGVLEGRDALPHILTLVPYKIS 182
 QY 182 AGEIKTKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKATLPTNAEKAVISAIDAD 241
 DB 183 AGEIKTKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKATLPTNAEKAVISAIDAD 242
 QY 242 TTYRIPILIRGQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 301
 DB 243 TTYRIPILIRGQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 302
 QY 302 DAYKSINELALHAGITRHKQVISTYIDSEITIEAGTAKLKNVDALIVPGGGERGVBEKI 361
 DB 303 DAYKSINELALHAGITRHKQVISTYIDSEITIEAGTAKLKNVDALIVPGGGERGVBEKI 362
 QY 362 STYRFARENKIPLYGICLGMQSAVIEFPAENVGLGASHTEFLPKSPHPVIGLITEMWDE 421
 DB 363 STYRFARENKIPLYGICLGMQSAVIEFPAENVGLGASHTEFLPKSPHPVIGLITEMWDE 422
 QY 422 AGEIVATRDSDSLGTMRLGAQCRKLKADSLAFOLYQKDVITERRRHRHRYEFNNQYLKLE 481
 DB 423 AGEIVATRDSDSLGTMRLGAQCRKLKADSLAFOLYQKDVITERRRHRHRYEFNNQYLKLE 482
 QY 482 AAGMKFSGKSLDGLVEITIELEHWPFLACQFHEPFTSTPRNGHALPSGFVBAAK 537
 DB 483 AAGMKFSGKSLDGLVEITIELEHWPFLACQFHEPFTSTPRNGHALPSGFVBAAK 538

RESULT 5
 H91083
 CTP synthetase [imported] - *Escherichia coli* (strain 0157:H7, substrain RIMD 0509952)

C;Species: *Escherichia coli*
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C;Accession: H91083
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasaawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: H91083
 A>Status: preliminary
 A:Molecule type: DNA
 A;Residues: 1-545 <HAY>
 A;Cross-references: GB:BA000907; PIDN:BA37063.1; PID:G13363111; GSPDB:GN00154
 A;Experimental source: strain 0157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: EC93640
 C;Superfamily: CTP synthase

Query Match 68.7%; Score 1916; DB 2; Length 545;
 Best Local Similarity 67.7%; Pred. No.2.5e-115;

Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

QY 2 TKEFIITGGVSSLSKGIAGASLSAILEDGKVTITKLDPIYNDPGTMSPOHGEVY 61
 DB 3 TTYIFVTGGVSSLSKGIAGASLSAILEDGKVTITKLDPIYNDPGTMSPOHGEVY 62
 QY 62 TEDGATDLDGHTERFRTKMTKKNFTTGOVYEQVLENERKDYLGATVQVPHITDE 121
 DB 63 TEDGATDLDGHTERFRTKMTKKNFTTGOVYEQVLENERKDYLGATVQVPHITDE 122
 QY 122 IKRRVYASAEKQVALIEVGTGDIESTPLETIRONGVLEGRDALPHILTLVPYKIS 181
 DB 123 IKRRVYASAEKQVALIEVGTGDIESTPLETIRONGVLEGRDALPHILTLVPYKIS 182
 QY 182 AGEIKTKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKATLPTNAEKAVISAIDAD 241
 DB 183 AGEIKTKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKATLPTNAEKAVISAIDAD 242
 QY 242 TTYRIPILIRGQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 301
 DB 243 TTYRIPILIRGQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 302
 QY 302 DAYKSINELALHAGITRHKQVISTYIDSEITIEAGTAKLKNVDALIVPGGGERGVBEKI 361
 DB 303 DAYKSINELALHAGITRHKQVISTYIDSEITIEAGTAKLKNVDALIVPGGGERGVBEKI 362
 QY 362 STYRFARENKIPLYGICLGMQSAVIEFPAENVGLGASHTEFLPKSPHPVIGLITEMWDE 421
 DB 363 STYRFARENKIPLYGICLGMQSAVIEFPAENVGLGASHTEFLPKSPHPVIGLITEMWDE 422
 QY 422 AGEIVATRDSDSLGTMRLGAQCRKLKADSLAFOLYQKDVITERRRHRHRYEFNNQYLKLE 481
 DB 423 AGEIVATRDSDSLGTMRLGAQCRKLKADSLAFOLYQKDVITERRRHRHRYEFNNQYLKLE 482
 QY 482 AAGMKFSGKSLDGLVEITIELEHWPFLACQFHEPFTSTPRNGHALPSGFVBAAK 537
 DB 483 AAGMKFSGKSLDGLVEITIELEHWPFLACQFHEPFTSTPRNGHALPSGFVBAAK 538

RESULT 6
 A85929
 CTP synthetase [imported] - *Escherichia coli* (strain 0157:H7, substrain EDL933)
 C;Species: *Escherichia coli*
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: A85929
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamotis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: A85929
 A>Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-545 <STO>
 A:Cross-references: GB:AE005174; MID:g12512252; PIDN:AA057893.1; GSPDB:GN00145; UMGPR:240
 A:Experimental source: strain 0157:H7, substrain EDL933
 C:Genetics:
 A:Gene: *pyrG*
 C:Superfamily: CTP synthase

Query Match 68.7%; Score 1916; DB 2; Length 545;
 Best Local Similarity 67.7%; Pred. No. 2.5e-115;
 Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

```

QY 2 TKFIFITGGVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSFGHGEVYV 61
DB 3 TNYIFVTGGVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSFGHGEVYV 62
QY 62 TEDGAEITDLDGHERFLKTKMTKNNFTTQGVYEQVLSNERKGDYLGATVQVIPHITDE 121
DB 63 TEDGAEITDLDGHERFLKTKMTKNNFTTQGVYEQVLSNERKGDYLGATVQVIPHITDE 122
QY 122 IKRRVYSAEGKDVALLVGGTVGDIISLPLETTIRQMGVGLGRDRALEFHLTLVPIYKS 181
DB 123 IKRRVYSAEGKDVALLVGGTVGDIISLPLETTIRQMGVGLGRDRALEFHLTLVPIYKS 182
QY 182 AGELTKRTPQSHVKEKRTIGIQPDLICRSEQIPASERKIALFTVAEKAVISALDAD 241
DB 183 SGVVKTKPQSHVKEKRTIGIQPDLICRSDRAVPAERAKIALFCNVEEKAVISLKDVD 242
QY 242 TTYRIPLLRQGLDLDVQDLRLDVPADLSAMEKVDGLTHTPDEVSAIYVGYVDHT 301
DB 243 SIYKIPGLKSGGLDLDYICRPSLCPENALAEWQVLYEESNPGEGYTIQMGKGYVLP 302
QY 302 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGEVKI 361
DB 303 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGEVKI 362
QY 362 STVRFARENKIPYLGICLGMQSAVIEFARNVGLBGAHSTFLPKSPHPVGLITEMDE 421
DB 363 TTRAFARENKIPYLGICLGMQSAVIEFARNVGLBGAHSTFLPKSPHPVGLITEMDE 422
QY 422 AGELVTRDESDLGSTMELGAOKRLKADSLAFOLYQKQVITERHRRHRYEFNNQYLKOLE 481
DB 423 NGNVEVRSKSDJGGTMRDLGAOKQLVDSLVRLQVYAPTVTERHRRHRYEFNNQYLKOLE 482
QY 482 AAGMKFSKSLDGLRVEILPEHPWFLACQHPHFTSTPRNGHALFSGFEVAAAK 537
DB 483 DAGIRAVAGSGDQVLEIIVPMHPVACQHPHFTSTPRNGHALFSGFEVAAAK 538

```

RESULT 7

AC0410
 CTP synthase (EC 6.3.4.2) [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC0410
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-1ariga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AC0410
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-545 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CA092607.1; PID:g15981303; GSPDB:GN00175
 C:Genetics:
 A:Gene: *pyrG*
 C:Superfamily: CTP synthase
 C:Keywords: ligase

Query Match 68.5%; Score 1909; DB 2; Length 545;
 Best Local Similarity 66.2%; Pred. No. 7.1e-115;
 Matches 356; Conservative 80; Mismatches 102; Indels 0; Gaps 0;

```

QY 2 TKFIFITGGVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSFGHGEVYV 61
DB 3 TNYIFVTGGVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSFGHGEVYV 62
QY 62 TEDGAEITDLDGHERFLKTKMTKNNFTTQGVYEQVLSNERKGDYLGATVQVIPHITDE 121
DB 63 TEDGAEITDLDGHERFLKTKMTKNNFTTQGVYEQVLSNERKGDYLGATVQVIPHITDE 122
QY 122 IKRRVYSAEGKDVALLVGGTVGDIISLPLETTIRQMGVGLGRDRALEFHLTLVPIYKS 181
DB 123 IKRRVYSAEGKDVALLVGGTVGDIISLPLETTIRQMGVGLGRDRALEFHLTLVPIYKS 182
QY 182 AGELTKRTPQSHVKEKRTIGIQPDLICRSEQIPASERKIALFTVAEKAVISALDAD 241
DB 183 SGVVKTKPQSHVKEKRTIGIQPDLICRSDRAVPAERAKIALFCNVEEKAVISLKDVD 242
QY 242 TTYRIPLLRQGLDLDVQDLRLDVPADLSAMEKVDGLTHTPDEVSAIYVGYVDHT 301
DB 243 SIYKIPGLKSGGLDLDYICRPSLCPENALAEWQVLYEESNPGEGYTIQMGKGYVLP 302
QY 302 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGEVKI 361
DB 303 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGEVKI 362
QY 362 STVRFARENKIPYLGICLGMQSAVIEFARNVGLBGAHSTFLPKSPHPVGLITEMDE 421
DB 363 TTRAFARENKIPYLGICLGMQSAVIEFARNVGLBGAHSTFLPKSPHPVGLITEMDE 422
QY 422 AGELVTRDESDLGSTMELGAOKRLKADSLAFOLYQKQVITERHRRHRYEFNNQYLKOLE 481
DB 423 NGNVEVRSKSDJGGTMRDLGAOKQLVDSLVRLQVYAPTVTERHRRHRYEFNNQYLKOLE 482
QY 482 AAGMKFSKSLDGLRVEILPEHPWFLACQHPHFTSTPRNGHALFSGFEVAAAKH 539
DB 483 AAGLAVAGRSADNKLVEILPEHPWFLACQHPHFTSTPRNGHALFSGFEVAAADYQ 540

```

RESULT 8

CP181
 CTP synthase (EC 6.3.4.2) - Haemophilus influenzae (strain Rd Kw20)
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 16-Jul-1999
 C:Accession: F64181
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kozlavage, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weisman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritschman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: F64181
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-545 <TIGR>
 A:Cross-references: GB:U37788; GB:I42073; MID:g1574629; PIDN:AA022733.1; PID:g1574630; 1
 C:Superfamily: CTP synthase
 C:Keywords: ligase; pyrimidine nucleotide biosynthesis

Query Match 68.5%; Score 1908.5; DB 2; Length 545;
 Best Local Similarity 67.5%; Pred. No. 7.6e-115;
 Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

```

QY 2 TKFIFITGGVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSFGHGEVYV 61
DB 3 TNYIFVTGGVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDPGTMSFGHGEVYV 62
QY 62 TEDGAEITDLDGHERFLKTKMTKNNFTTQGVYEQVLSNERKGDYLGATVQVIPHITDE 121
DB 63 TEDGAEITDLDGHERFLKTKMTKNNFTTQGVYEQVLSNERKGDYLGATVQVIPHITDE 122
QY 122 IKRRVYSAEGKDVALLVGGTVGDIISLPLETTIRQMGVGLGRDRALEFHLTLVPIYKS 181

```

Db 123 IKDVIAGAGQGHVIVVGVGVGDIIESLPLEALRQIAVQNGREHTLFMELTLVPIPT 182
QY 182 AGELTKPTQHSYKELRTIGIOPDILICRSBQIPASERRKIALFTNVAEKAVISALDA 241
Db 183 ABEVTKPTQHSYKELRTIGIOPDILICRSBQIPASERRKIALFTNVAEKAVISALDA 242
QY 242 TIYRIPLLEBQGLDLDVQDLRLVPAADLSAMEKVDGLTHPPDEVSIALVGRKYV 301
Db 243 SIYQIPALLKSGQDLDFVCEERFLTCPEADLTWEDEVLYKQANPVEEVIYIGVQKTELP 302
QY 302 DAVKSLNEALIHAGIHTFRKVOISYIDSETEBAEGTAKIKNVDALVPGFSGERGVEGI 361
Db 303 DAVKSLNEALIHAGIHTFRKVOISYIDSETEBAEGTAKIKNVDALVPGFSGERGVEGI 362
QY 362 STYFAREKNIPTLGIICAMQSAVIEFARNVVGLEGASHTETLPKSPHPVIGLITEMDE 421
Db 363 RTQVAREKNIPTLGIICAMQSAVIEFARNVVGLEGASHTETLPKSPHPVIGLITEMDE 422
QY 422 AGEVLTFRDSDIGGTMRLGAQCKRLKADSLAFOLYQKDVITERHRRHRYEPNNYKOLE 481
Db 423 EGVTEVTRDSDIGGTMRLGAQCKRLKADSLAFOLYQKDVITERHRRHRYEPNNYKOLE 482
QY 482 AAGMKFSKSLDGRVLEIIEPEHPWFLACOPHPETSTPRNGHALFSGVEAA-AKHK 539
Db 483 KGLKATVGLSADKQVLEIIEPEHPWFLACOPHPETSTPRNGHALFSGVEAA-AKHK 541

RESULT 9

C82700
CTP synthetase XF1268 (imported) - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: C82700
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
Nucleotide 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: C82700
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <SIM>
A:Cross-References: GB:AE003962; GB:AE003849; NID:G9106270; PIDN:AAF84097.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Ascencio, M.; Alvarenga, R.; A-
Briomes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H-
as-Neco, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm-
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig-
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E-
A:Authors: Martins, E.M.F.; Matsubara, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa-
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., M.A.; da Silva
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1268
C:Superfamily: CTP synthase

Query Match 68.3%; Score 1905; DB 2; Length 554;
Best Local Similarity 66.1%; Pred. No. 1.3e-114;
Matches 361; Conservative 79; Mismatches 102; Indels 4; Gaps 2;
QY 1 MKKFIITGVSVSSIGKIAASSLAILEDRLKVTITLDPYINVDGTMSPFOHGEVF 60
Db 1 MKKFIITGVSVSSIGKIAASSLAILEDRLKVTITLDPYINVDGTMSPFOHGEVF 60
QY 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKQDYLGATVQVPHITD 120
Db 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKQDYLGATVQVPHITD 120
QY 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKQDYLGATVQVPHITD 120
Db 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKQDYLGATVQVPHITD 120

QY 121 EIKRNVESAEGKVALTEGVGVGDIIESLPLETIROMGYELGRDALFHTLTVPIK 180
Db 121 EIKRNVESAEGKVALTEGVGVGDIIESLPLETIROMGYELGRDALFHTLTVPIK 180
QY 181 SAGELTKPTQHSYKELRTIGIOPDILICRSBQIPASERRKIALFTNVAEKAVISALDA 240
Db 181 SAGELTKPTQHSYKELRTIGIOPDILICRSBQIPASERRKIALFTNVAEKAVISALDA 240
QY 241 PTYRIPLLEBQGLDLDVQDLRLVPAADLSAMEKVDGLTHPPDEVSIALVGRKYV 298
Db 241 PTYRIPLLEBQGLDLDVQDLRLVPAADLSAMEKVDGLTHPPDEVSIALVGRKYV 298
QY 298 DHTDAKSLNEALIHAGIHTFRKVOISYIDSETEBAEGTAKIKNVDALVPGFSGERGVE 358
Db 298 DHTDAKSLNEALIHAGIHTFRKVOISYIDSETEBAEGTAKIKNVDALVPGFSGERGVE 358
QY 358 GKI STYFAREKNIPTLGIICAMQSAVIEFARNVVGLEGASHTETLPKSPHPVIGLITEM 418
Db 358 GKI STYFAREKNIPTLGIICAMQSAVIEFARNVVGLEGASHTETLPKSPHPVIGLITEM 418
QY 418 AGEVLTFRDSDIGGTMRLGAQCKRLKADSLAFOLYQKDVITERHRRHRYEPNNYKOLE 481
Db 418 AGEVLTFRDSDIGGTMRLGAQCKRLKADSLAFOLYQKDVITERHRRHRYEPNNYKOLE 481
QY 481 RTTGEVTRDSDIGGTMRLGAQCKRLKADSLAFOLYQKDVITERHRRHRYEPNNYKOLE 536
Db 481 RTTGEVTRDSDIGGTMRLGAQCKRLKADSLAFOLYQKDVITERHRRHRYEPNNYKOLE 536
QY 536 KHKTOG 542
Db 542 KHKTOG 546

RESULT 10

F81798
CTP synthase (EC 6.3.4.2) NMA1742 (imported) - Neisseria meningitidis (strain Z2491 serog-
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: F81798
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nucleotide 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; PMID:20222556; PMID:10761919
A:Accession: F81798
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <PAR>
A:Cross-References: GB:AL162757; GB:AL157959; NID:G7380371; PIDN:CA884970.1; PID:G7380384
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: pyg; NMA1742
C:Superfamily: CTP synthase
C:Keywords: ligase

Query Match 67.9%; Score 1892.5; DB 2; Length 544;
Best Local Similarity 66.1%; Pred. No. 8.1e-114;
Matches 358; Conservative 80; Mismatches 101; Indels 3; Gaps 2;
QY 1 MKKFIITGVSVSSIGKIAASSLAILEDRLKVTITLDPYINVDGTMSPFOHGEVF 60
Db 1 MKKFIITGVSVSSIGKIAASSLAILEDRLKVTITLDPYINVDGTMSPFOHGEVF 60
QY 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKQDYLGATVQVPHITD 120
Db 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKQDYLGATVQVPHITD 120
QY 121 EIKRNVESAEGKVALTEGVGVGDIIESLPLETIROMGYELGRDALFHTLTVPIK 180
Db 121 EIKRNVESAEGKVALTEGVGVGDIIESLPLETIROMGYELGRDALFHTLTVPIK 180
QY 181 SAGELTKPTQHSYKELRTIGIOPDILICRSBQIPASERRKIALFTNVAEKAVISALDA 240
Db 181 SAGELTKPTQHSYKELRTIGIOPDILICRSBQIPASERRKIALFTNVAEKAVISALDA 240


```

      361 KLLSTIYAAEENHPIYGLICGMQIAIEFAQNVVKGKANSSTFEDQCKYPIIDLKNRP 420
      420 DEAGELVTEDEDS-DIGTWRLLGAQKCRLLKADSLAQOLYOKDVTERRHRYEFNNQYLK 478
      421 NNSSKVNNKINENINIGCTMRLOSQPCCKLSANLSKRLNNOELLIRHRYEVNVLKPK 480
      479 QLEAAGKSGSGSLDGRVLEIILPEHPWFLACQHPHPTSTPRNGHALPSGFVEAAAKH 538
      481 KLEAAGLVGVRSQKNNVVEIILSNHPWFLACQHPHPTSTPRDCHPLFIDFKSAGKH 540
      539 K 539
      541 K 541

RESULT 13
CTP synthetase ctra (imported) - Bacillus halodurans (strain C-125)
H84123
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #ext_change 15-Jun-2001
C:Accession: H84123
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MID:20512562; PMID:11058132
A:Accession: H84123
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-532 <STO>
A:Cross-references: GB:AP001520; GB:BA000004; MID:g10176401; PIDN:BA807511.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: CTP synthase

Query Match      59.0%; Score 1645.5; DB 2; Length 532;
Best Local Similarity 57.7%; Pred. No. 5,5e-98;
Matches 311; Conservative 90; Mismatches 125; Indels 13; Gaps 3;

      2 TKRIFITGGVSSLGKGIASSLAAILDEBGLKVTITKLDPIYINVDPGTMSPPQGEVY 61
      3 TKYIFITGGVSSLGKGIASSLAAILDEBGLKVTITKLDPIYINVDPGTMSPPQGEVY 62
      62 TEDGAEITDLDGHERPFLKTTMKKONFTTGGVYEVOLNTERKADYLGATVQVPIHTD 121
      63 TIDGAEITDLDGHERPFLKTTMKKONFTTGGVYEVOLNTERKADYLGATVQVPIHTD 122
      122 IKRVSYSAAE--GKVALIEVGSTVGDIESLPLETTIRQKGYELGRDALFIHLTVPIY 179
      123 IKRVSYSAAE--GKVALIEVGSTVGDIESLPLETTIRQKGYELGRDALFIHLTVPIY 182
      180 KSAQELKTPYQHSVKELEPFIQIPDILCRSEQPIPAERKRIAFVTVAEKAIVSAID 239
      183 AAGGKSKPTQHSVKELEPFIQIPDILCRSEQPIPAERKRIAFVTVAEKAIVSAID 242
      240 ADITVYIPILLRQGLDDLVVDQLRDVPAADLSAMEKVVDGLTHPTDEVSIAIVGYVD 299
      243 ADITVYIPILLRQGLDDLVVDQLRDVPAADLSAMEKVVDGLTHPTDEVSIAIVGYVD 302
      300 HTDPAVLSNEALIHAGIHTHRKQVQISYIDSETEAAGTA-KLKNVDAIIVPGFGRGVE 358
      303 LPAVYLSNEALIHAGIHTHRKQVQISYIDSETEAAGTA-KLKNVDAIIVPGFGRGVE 362
      359 GKISTVPAEENKIPYIGICGMQSAVIERARVVGLEGASSTFEPKSPHPVIGLITEM 418
      363 GKIEATVPAEENKIPYIGICGMQSAVIERARVVGLEGASSTFEPKSPHPVIGLITEM 422
      419 MDAGSLVTRDESDIGTWRLLGAQKCRLLKADSLAQOLYOKDVTERRHRYEFNNQYLK 478
      423 KDV-----EDWGTIRLGLYPCCKKNGTLAGSANTDQVYERHRYEFNNQYRE 472
      479 QLEAAGKSGSGSLDGRVLEIILPEHPWFLACQHPHPTSTPRNGHALPSGFVEAAAK 537
  
```

```

      473 QLEAAGKSGSGSLDGRVLEIILPEHPWFLACQHPHPTSTPRNGHALPSGFVEAAAK 531
      531 K 531

RESULT 14
CTP synthetase - Aquifex aeolicus
P70415
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #ext_change 16-Jul-1999
C:Accession: P70415
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MID:98196666; PMID:9537320
A:Accession: P70415
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-531 <AOF>
A:Cross-references: GB:AB000735; MID:g2983749; PIDN:AAC07314.1; PID:g2983754; GB:AB00065;
A:Experimental source: strain VFS
C:Genetics:
C:Superfamily: CTP synthase

Query Match      58.8%; Score 1639; DB 2; Length 531;
Best Local Similarity 56.9%; Pred. No. 1.4e-97;
Matches 309; Conservative 94; Mismatches 128; Indels 12; Gaps 2;

      1 MKRIFITGGVSSLGKGIASSLAAILDEBGLKVTITKLDPIYINVDPGTMSPPQGEVY 60
      1 MKYIFITGGVSSLGKGIASSLAAILDEBGLKVTITKLDPIYINVDPGTMSPPQGEVY 60
      61 VTEDEAEITDLDGHERPFLKTTMKKONFTTGGVYEVOLNTERKADYLGATVQVPIHTD 120
      62 VTEDEAEITDLDGHERPFLKTTMKKONFTTGGVYEVOLNTERKADYLGATVQVPIHTD 120
      121 EIKRVSYSABQVALIEVGSTVGDIESLPLETTIRQKGYELGRDALFIHLTVPIYK 180
      122 EIKRVSYSABQVALIEVGSTVGDIESLPLETTIRQKGYELGRDALFIHLTVPIYK 180
      181 SAGELKTPYQHSVKELEPFIQIPDILCRSEQPIPAERKRIAFVTVAEKAIVSAID 240
      184 AAGELKTPYQHSVKELEPFIQIPDILCRSEQPIPAERKRIAFVTVAEKAIVSAID 240
      241 DITVYIPILLRQGLDDLVVDQLRDVPAADLSAMEKVVDGLTHPTDEVSIAIVGYVD 300
      242 DITVYIPILLRQGLDDLVVDQLRDVPAADLSAMEKVVDGLTHPTDEVSIAIVGYVD 300
      243 DITVYIPILLRQGLDDLVVDQLRDVPAADLSAMEKVVDGLTHPTDEVSIAIVGYVD 300
      301 TDAYSLSNEALIHAGIHTHRKQVQISYIDSETEAAGTA-KLKNVDAIIVPGFGRGVE 360
      304 KDSYSVTEALIHAGIHTHRKQVQISYIDSETEAAGTA-KLKNVDAIIVPGFGRGVE 360
      361 ISTVPAEENKIPYIGICGMQSAVIERARVVGLEGASSTFEPKSPHPVIGLITEM 420
      364 ISTVPAEENKIPYIGICGMQSAVIERARVVGLEGASSTFEPKSPHPVIGLITEM 420
      421 EAGELVTEDESDIGTWRLLGAQKCRLLKADSLAQOLYOKDVTERRHRYEFNNQYLK 480
      424 EAGELVTEDESDIGTWRLLGAQKCRLLKADSLAQOLYOKDVTERRHRYEFNNQYLK 480
      481 EAGELVTEDESDIGTWRLLGAQKCRLLKADSLAQOLYOKDVTERRHRYEFNNQYLK 480
      484 EAGELVTEDESDIGTWRLLGAQKCRLLKADSLAQOLYOKDVTERRHRYEFNNQYLK 480
      541 QCF 543
      529 KTF 531

RESULT 15
CTP synthase (EC 6.3.4.2) - Bacillus subtilis
  
```

N:Alternate names: CTP-synthetase; UTP-ammونيا ligase
 C:Species: *Bacillus subtilis*
 C:Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text change 16-Jun-2000
 C:Accession: A32354; S55423; C69610
 R:Trach, K.; Chapman, J.M.; Pigott, P.; Lecocq, D.; Hoch, J.A.
 J. Bacteriol. 170, 4194-4208, 1988
 A:Title: Complete sequence and transcriptional analysis of the spoOF region of the *Bacillus subtilis* chromosome
 A:Accession: A32354
 A:Reference number: A31883; PMID:88314920; PMID:2457578
 A:Molecule type: DNA
 A:Residues: 1-535 <RNA>
 A:Cross-references: GB:M22039; NID:G460310; PIDN:AA16801.1; PID:G143597
 R:Laaser, P.; Danchin, A.
 Submitted to the EMBL Data Library, May 1995
 A:Description: Cloning and sequencing of the *Bacillus subtilis* chromosomal region from 3
 A:Reference number: S55414
 A:Accession: S55423
 A:Molecule type: DNA
 A:Residues: 1-535 <GLA>
 A:Cross-references: EMBL:249782; NID:G853752; PIDN:CA89870.1; PID:G853762
 R:Kumst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bettec
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, B.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei
 Y, M.; Ogawa, K.; Ogihara, A.; Oudea, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schreier, R.; Scelfone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 akouch, M.; Tanakoshi, A.; Tanaka, T.; Tepsstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A65580; PMID:98044033; PMID:9384377
 A:Accession: C69610
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-535 <RNA>
 A:Cross-references: GB:299123; GB:A1009126; NID:G2636240; PIDN:CA15743.1; PID:G2636252
 A:Experimental source: strain 168
 C:Comment: This enzyme is a glutamine amidotransferase that catalyzes the terminal react
 C:Genetics:
 A:Gene: ctpA; pyrG
 A:Map position: 37 min
 C:Superfamily: CTP synthase
 C:Keywords: ligase; pyrimidine nucleotide biosynthesis

Query Match 58.2%; Score 1622.5; DB 1; Length 535;
 Best Local Similarity 57.0%; Pred. No. 1.7e-96;
 Matches 310; Conservative 94; Mismatches 127; Indels 13; Gaps 3;

QY 1 MTKFIIITGVSLSKXGIAASLAILLEDRGLKVTITLDPYINVDPTGMSPPRQGRVP 60
 DB 1 MTKYIFVTGVSLSKXGIAASLAILLEDRGLKVTITLDPYINVDPTGMSPPRQGRVP 60
 QY 61 VTEDGAEITDLDGHERFPLTKMTKKNFTTQVYEVLNERKGDYLGATVQVPHITD 120
 DB 61 VTDDGAEITDLDGHERFPLTKMTKKNFTTQVYEVLNERKGDYLGATVQVPHITN 120
 QY 121 EIKRYYESAE--GKDVALLIEVGTYGDIESTPLETTFQMGVLEGDRBALFHLTLVPY 178
 DB 121 EIKDRYYRAGKSTNADVLTTEIGTVGDIESTPLETTFQMGVLEGDRBALFHLTLVPY 180
 QY 179 IKSAGELKTPQHSYKELRTIGIQDILICSEOPIPASERKIALFTNVAEKAVISA 238
 DB 181 IKAAGELKTPQHSYKELRTIGIQDILICSEOPIPASERKIALFTNVAEKAVISAI 240
 QY 239 DADTIRIRLLRBOGLDVLVDQLRDVPAADLSAMEKVDGLTHPTDEVSIAIVSKY 298
 DB 241 DADNYSIFELQKQGLDVLVDQLRDVPAADLSAMEKVDGLTHPTDEVSIAIVSKY 300

QY 299 DETDAYSLNEALIHAGIHTKHKVOISYIDSETIEAGSTAXL-KNDAILVPGGPRGCV 357
 DB 301 ELPPAYISVSESLRHAGYAFPTDVKVKNIAEEYENNAIAELTSGTGLIIVPGGPRGCV 360
 QY 358 ECKSTYREFANENKIPYIGICLGWASIVBARVVLGQHSFELPKSPHPIGLITE 417
 DB 361 ECKIVATYKARENKIPYIGICLGWASIVBARVVLGQHSFELPKSPHPIGLITE 420
 QY 418 MWDSAGELVTRDESDIDGTMRGAKCRKLKAUSLAEQLYQKDYITERHRRHYEFNNQYL 477
 DB 421 QKDY-----EDIGTFLGLYPCCKLEGTAKAEVYQDEVVYRHRHRYEFNNIEFR 470
 QY 478 KQLNAGKRSKSGSLDGRVHITLPEHPPTLACQHFPTSTPRNKAHAFSGFVEAAX 537
 DB 471 QQMEEGVFESGTSPPDGLVETITLKHDPFWVASQFHPFRSPTPPDPLKFGTASVE 530
 QY 538 HKTQ 541
 DB 531 AANQ 534

Search completed: January 29, 2004, 15:55:33
 Job time : 17.0063 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 10.244 Seconds

(without alignment) 2497.314 Million cell updates/sec

Title: US-09-941-947A-16

Sequence: 1 MTKFIFITGVGVSLGKGLA.....HALFGFVFAAKKTKGTGA 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	72.0	542	1	PYRG_PSEAE
2	1948	69.9	564	1	PYRG_NITRU
3	1946	68.8	544	1	PYRG_SALTU
4	1938	69.5	554	1	PYRG_XANCP
5	1937	69.5	546	1	PYRG_VIBPA
6	1931	69.3	554	1	PYRG_XANAC
7	1925	69.0	545	1	PYRG_VIBVU
8	1924	69.0	544	1	PYRG_VIBCH
9	1918	68.8	542	1	PYRG_PASRU
10	1916	68.7	544	1	PYRG_ECOLI
11	1909	68.5	544	1	PYRG_YERPE
12	1908.5	68.5	545	1	PYRG_HAEIN
13	1905	68.3	554	1	PYRG_XYLFA
14	1892.5	67.9	544	1	PYRG_NEIMA
15	1888.5	67.7	544	1	PYRG_NEIMA
16	1822	65.4	554	1	PYRG_RALSO
17	1737	62.3	545	1	PYRG_BUCAL
18	1729.5	62.0	553	1	PYRG_BUCAP
19	1685.5	60.5	544	1	PYRG_BUCAP
20	1662	59.6	552	1	PYRG_WIGAR
21	1645.5	58.0	532	1	PYRG_BACED
22	1640.5	58.8	537	1	PYRG_THERN
23	1639	58.8	531	1	PYRG_AOURA
24	1622.5	58.2	535	1	PYRG_BACSU
25	1615.5	57.9	535	1	PYRG_CLOAB
26	1603.5	57.4	552	1	PYRG_SYNT3
27	1600.5	57.4	532	1	PYRG_LISIN
28	1600.5	57.4	530	1	PYRG_LISNO
29	1595	57.2	550	1	PYRG_CAUOR
30	1592	57.1	542	1	PYRG_BRUNE
31	1586	56.9	542	1	PYRG_RHINE
32	1583.5	56.8	535	1	PYRG_CLOPE
33	1580.5	56.7	546	1	PYRG_SYNP7

34	1576	56.5	536	1	PYRG_STAMX
35	1574.5	56.5	545	1	PYRG_ANASP
36	1563.5	56.1	540	1	PYRG_METUA
37	1562	56.0	535	1	PYRG_STERN
38	1562	56.0	542	1	PYRG_AGRFS
39	1555.5	55.8	537	1	PYRG_PYRAB
40	1555	55.8	542	1	PYRG_RHILU
41	1552.5	55.7	535	1	PYRG_METKA
42	1542	55.3	544	1	PYRG_AZORR
43	1540.5	55.3	537	1	PYRG_PYRPU
44	1538.5	55.2	537	1	PYRG_PYRHO
45	1507	54.1	543	1	PYRG_CAMJR

ALIGNMENTS

RESULT 1
ID PYRG_PSEAE STANDARD; PRT; 542 AA.
AC Q9HXZ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN PYRG OR PA3637.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Gardner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Mu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
RL -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
either L-glutamine or ammonia as the source of nitrogen (By
similarity).
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
is the substrate. Inhibited by CTP (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP, third
(last) step.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SIMILARITY: Belongs to the CTP synthase family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).
CC -----
CC EMBL, AB004783; AB007025.1; -;
CC PIR, B83192; B83192.
CC HAMAP, MF_01227; -1.
CC InterPro, IPR000991; GATase_1.
CC InterPro, IPR004468; Pyrg.
CC Pfam, PF00117; GATase_1.
CC TIGRfam, TIGR00337; Pyrg.1.
DR PROSITE, PS00442; GATASE_TYP_1; 1.
DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.


```

Db      361  IMAISVANHRIPIYIGICLGMQLAVIESRNRLOENNAHSTEFEDPTPYVGLITTEWRD 420
Qy      421  EAGELVTEDESDLOGTMRIGAKORCLADSLAPLOYKDVITEHRRHRYFNNOYKOL 480
Db      421  RCGREKESAGQDDLSGTMRLAGQCLKPRYTHAKTYGADKYIEHRRHRYFNNAFFIOL 480
Qy      481  EAGKRFSGKSLDGLVLEIIEIP--EHPWFLACQHPPEFTSPNGHALFSGVEAA 535
Db      481  RQAGKHISGLSABEDLCMIETLPEQEHHPVFAQCQHPPEFTSPNGHPLFKSYIOAA 537

RESULT 3
PYRG_SALTY STANDARD; PRT; 544 AA.
AC      O8XEN5;
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
OS      PYRG OR STM2953 OR STY3082 OR T2854.
OC      Salmonella typhimurium, and
OC      Salmonella typhi.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Salmonella.
OC      NCBI_TaxId=602, 601;
RN      (1)
RN      SEQUENCE FROM N.A.
RP      SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX      MEDLINE=21534948; PubMed=11677609;
RA      McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA      Courtney L., Porcullik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA      Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA      Ryan S., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA      Waterston R., Wilson R.K.;
RT      "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT      LT2."
RT      Nature 413:852-856(2001).
RN      (2)
RN      SEQUENCE FROM N.A.
RP      SPECIES=S.typhi; STRAIN=CT18;
RX      MEDLINE=21534947; PubMed=11677608;
RA      Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA      Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA      Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA      Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA      Pelletier T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA      Krogan A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA      Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA      Whitehead S., Barrett B.G.;
RT      "Complete genome sequence of a multiple drug resistant Salmonella
RT      enterica serovar Typhi CT18."
RT      Nature 413:848-852(2001).
RN      (3)
RN      SEQUENCE FROM N.A.
RP      SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX      MEDLINE=22531367; PubMed=12644504;
RA      Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA      Burland V., Kodymani V., Schwartz D.C., Blattner F.R.;
RT      "Comparative genomes of Salmonella enterica serovar Typhi strains Ty2
RT      and CT18."
RT      J. Bacteriol. 185:2330-2337(2003).
RN      (4)
RN      FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
RN      either L-glutamine or ammonia as the source of nitrogen (by
RN      similarity).
RN      CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
RN      -1- ENZYME REGULATION: Allosterically activated by GMP, when glutamine
RN      is the substrate. Inhibited by CTP (by similarity).
RN      -1- PATHWAY: Pyrimidine biosynthesis; conversion of UTP to CTP, third
RN      (last) step.
RN      -1- SUBUNIT: Homotrimer (by similarity).
RN      -1- SIMILARITY: Belongs to the CTP synthase family.
RN      -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

```

```

CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AEO08835; AAL21833.1; -
CC      EMBL; AL627276; CAD06059.1; -
CC      EMBL; AEO16843; AA070411.1; -
CC      StyGene; SG72727; P1.
CC      DR      HAVAP; MF_01227; -; 1.
CC      DR      InterPro; IPR000991; GATase_1.
CC      DR      Pfam; PF00117; GATase; 1.
CC      DR      PROSITE; PS00442; GATASE TYPE I; 1.
CC      KW      Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
CC      Complete proteome.
CC      FT      INIT MET 0
CC      FT      DOMAIN 0
CC      FT      DOMAIN 1 299
CC      FT      ACT_SITE 300 544
CC      FT      ACT_SITE 378 514
CC      FT      ACT_SITE 514 516
CC      FT      ACT_SITE 516 516
CC      SQ      SEQUENCE 544 AA; 5990 MW; BABCA85F342EAB8 CRC64;

Query Match 69.8%; Score 1946; DB 1; Length 544;
Best local similarity 68.2%; Pred. No. 1,4e-114;
Matches 367; Conservative 73; Mismatches 98; Indels 0; Gaps 0;

Qy      2  TKRFIFGVVSSIGKRIASLSAILEDGKLTIRKLPYINVDPGTSPPGHGVFV 61
Qy      2  TNYIFVGVVSSIGKRIASLSAILEDGKLTIRKLPYINVDPGTSPPGHGVFV 61
Db      62  TEDGAEITDLIGHYERLKTMTKKNFTTQVYVQVLANERKDYLGATVQVIEPHITDE 121
Db      62  TEDGAEITDLIGHYERLKTMTKKNFTTQVYVQVLANERKDYLGATVQVIEPHITNA 121
Qy      122  IKRNVESABQKVALLEVGVGVVDIESLPLETRIRMGVYELGDRALFHLTVPIYKS 181
Db      122  IKRNVESABQKVALLEVGVGVVDIESLPLETRIRMGVYELGDRALFHLTVPIYLA 181
Qy      182  AGEFKTPQYHVAELTIGIOPDILICRSEOPTPASEKRIALFTVNAKAVISAIDAD 241
Db      182  AGEFKTPQYHVAELTIGIOPDILICRSDRAVPAKRIALFCVVPKAVISMCDVD 241
Qy      242  TTYIPLILBQGLIDLVDOQLRDVPAOLSAKENVVDGLTHFTDVSIAIVKRYVDHT 301
Db      242  TTYIPLILBQGLIDLVDOQLRDVPAOLSAKENVVDGLTHFTDVSIAIVKRYVDHT 301
Qy      242  SIYRIPGLKSGQGLIDVYICRFSINCEPANTSEBQVYIEBANPAGVTTIGMVKYIELP 301
Db      242  SIYRIPGLKSGQGLIDVYICRFSINCEPANTSEBQVYIEBANPAGVTTIGMVKYIELP 301
Qy      302  DAYSLSNALIHAGIRHRYKVOISYIDSETIEAGTKLKNVDAILVPGSGRGEVBEKI 361
Db      302  DAYSLSNALIHAGIRHRYKVOISYIDSETIEAGTKLKNVDAILVPGSGRGEVBEKI 361
Qy      362  STVAPARENKIPIYIGICLQASAVIEPARVVGLEGASTIEFLPKSPHPVIGLITTEWRD 421
Db      362  STVAPARENKIPIYIGICLQASAVIEPARVVGLEGASTIEFLPKSPHPVIGLITTEWRD 421
Qy      422  ATAYARARENKIPIYIGICLQASAVIEPARVVGLEGASTIEFLPKSPHPVIGLITTEWRD 481
Db      422  ATAYARARENKIPIYIGICLQASAVIEPARVVGLEGASTIEFLPKSPHPVIGLITTEWRD 481
Qy      482  DGNVTVSRKSDLSGTMRLAGQCLKPRYTHAKTYGADKYIEHRRHRYFNNAFFIOL 539
Db      482  DGNVTVSRKSDLSGTMRLAGQCLKPRYTHAKTYGADKYIEHRRHRYFNNAFFIOL 539
Qy      482  AAGKRFSGKSLDGLVLEIIEIP--EHPWFLACQHPPEFTSPNGHALFSGVEAAAKK 539
Db      482  AAGKRFSGKSLDGLVLEIIEIP--EHPWFLACQHPPEFTSPNGHALFSGVEAAAKK 539

RESULT 4
PYRG_XANCP STANDARD; PRT; 554 AA.
AC      Q8P926;
DT      28-FEB-2003 (Rel. 41, Created)

```


DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 GN PYRG OR XCC1697.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xanthomonas.
 NX NCBI_TaxID=340;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA de Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Burlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.S.A.,
 RA Camarotte G., Camarvan F., Cardoso J., Chamberg F., Chaplin L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Fontignier E.F., Franco M.C., Greggio C.C., Guber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meladas J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
 RA Spicola L.A.F., Takita M.A., Tamura R.E., Teixeira R.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Serubaj J.C., Kitejima J.P.,
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RT Nature 417:459-463(2002).
 RL -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 CC either L-glutamine or ammonia as the source of nitrogen (3y
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 CC is the substrate. Inhibited by CTP (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP, third
 CC (last) step.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SIMILARITY: Belongs to the CTP synthase family.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AEO12271; AAM40991.1; ALT_INIT.
 CC HAMAP: MF_01227; -1.
 CC InterPro: IPR000991; GATase_1.
 CC InterPro: IPR004468; Pyrg.
 CC Pfam: PF00117; GATase; 1.
 CC TIGRfam: TIGR00337; Pyrg; 1.
 CC PROSITE: PS00442; GATAS_FTPB; 1.
 CC Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
 CC Complete proteome.
 CC
 CC DOMAIN 1 301 AMINATOR DOMAIN.
 CC FT DOMAIN 302 554 GLUTAMINE AMIDOTRANSFERASE.
 CC FT ACT SITE 380 380 GATASE (BY SIMILARITY).
 CC FT ACT SITE 518 518 GATASE (BY SIMILARITY).
 CC FT ACT SITE 520 520 GATASE (BY SIMILARITY).
 CC FT ACT SITE 554 AA; 61514 MW; A2D70CA2DBDF1AEO CR664;
 CC SEQUENCE
 CC
 CC Query Match 69.5%; Score 1938; DB 1; Length 554;
 CC Best Local Similarity 67.2%; Pred. No. 4,4e-114;
 CC Matches 367; Conservative 80; Mismatches 95; Indels 4; Gaps 2;
 CC
 CC QY 1 MYKFIITGGVSVSLKGIKIAASLAAILBEDGKLTITTTKLDPIYINDGCTMSPFGGEVF 60
 CC DB 1 MTPFLIVTGGVSVSLKGIKIAASLAAILBEDGKLTITTTKLDPIYINDGCTMSPFGGEVF 60

QY 61 VTDEGAEITDIDLGHTESFLKTTMTKKNPTTGGVYHVOVLNENKGYLCAATVYTHITD 120
 DB 61 VTDDGAEITDIDLGHTESFLKTTMTKKNPTTGGVYHVOVLNENKGYLCAATVYTHITD 120
 QY 121 EIKRRVYSAEGKDVALLIEVGATVGDIESLPFETTRQMGVEKARDALEFHTLTVYIK 180
 DB 121 EIKRCIDEATGSPDVALLIEVGATVGDIESLPFETTRQMGVEKARDALEFHTLTVYIK 180
 QY 181 SAGELKTPQHSVKEIRTIQIOPDILICRSEQPIPAERKRTALFTNVAEKVISAIDA 240
 DB 181 AAGELKTPQHSVKEIRTIQIOPDILICRSEQPAVDPSEKRTALFTNVSERAVISCPDI 240
 QY 241 PTYIRIPIALLREGDLDVYDQRL--DYPAAALSAEKVVDGLTPTDEVSIAIVGKY 298
 DB 241 DVLGMPLELRQGLDELDVDFGLKRVAAADSEBAVADVAKHPLDEVITIAVAKIV 300
 QY 299 DHTDAVKSILNHALIHAGIHTRHVKVQISYIDSETIEAGTAKLNVAAILVPGGGERGVE 358
 DB 301 DHQDAVSVVAAALHGGIARQRTKVNLTWBAQDLESDMAALQIDGILVPGGGERGFE 360
 QY 359 GKISTVFAEKNTPIYIGICMGQSAITERANVVGEGHSTFPLKSPHPIGLITTEW 418
 DB 361 GKVGTSKYABHHKVPYIGICMGQAAVADVARRVADIDAA NSTENDROSPHPVIGITTEW 420
 QY 419 MDZAGELVTRDESDLDGTVRLAQKCRKLKADSLAPQLYOKDVITERRHRYEFNNQYLX 478
 DB 421 RTATGVEKRDENSDLDGTVRLAQKCRKLKADSLAPQLYOKDVITERRHRYEFNNQYRT 480
 QY 479 QLEPAKMGSGKSIDGLVETILELP--HFWFLACQHPERTSPRNGALFSGFVAAA 536
 DB 481 QLEBAGLVIGKSMDDTLVWVELPRDTHPMFLACQAHPEFLTPRGGHPLFIFVAAAR 540
 QY 537 KHKTOG 542
 DB 541 EKKAGG 546
 CC
 CC RESULT 5
 CC PRG_VIRBA STANDARD; PRT; 546 AA.
 CC ID PRG_VIRBA
 CC AC 0871B9;
 CC DT 15-SEP-2003 (Rel. 42, Created)
 CC DT 15-SEP-2003 (Rel. 42, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 CC GN PYRG OR VP2562.
 CC OS Vibrio parahaemolyticus.
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC CC Vibrionaceae; Vibrio.
 CC NX NCBI_TaxID=670;
 CC RN
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=RIMD 221063 / Serotype O3:K6;
 CC RX MEDLINE=22508454; PubMed=12620739;
 CC RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 CC RA Iijima Y., Naito M., Nakano M., Yamashita A., Kobota Y., Kimura S.,
 CC RA Yasunaga T., Honda T., Shinagawa H., Hattori K., Iida T.,
 CC RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 CC distinct from that of V. cholerae.";
 CC RL Lancet 361:743-749(2003).
 CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 CC either L-glutamine or ammonia as the source of nitrogen (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 CC is the substrate. Inhibited by CTP (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP, third
 CC (last) step.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SIMILARITY: Belongs to the CTP synthase family.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: APO05082; BAC60825.1; -

DR HAMAP: MF_01227; -

DR PROSITE: PS00442; GATASE_TYPE_1; 1.

KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.

FT DOMAIN 1 300 AMINATOR DOMAIN
FT ACT_SITE 301 546 GUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 379 379 GATASE (BY SIMILARITY).
FT ACT_SITE 515 515 GATASE (BY SIMILARITY).
FT ACT_SITE 517 517 GATASE (BY SIMILARITY).
SQ SEQUENCE 546 AA; 60126 MW; E33C935E45B89E4 CRC64;

Query Match 69.5%; Score 1937; DB 1; Length 546;
Best Local Similarity 68.2%; Pred. No. 5e-116; Mismatches 96; Gaps 0;

Matches 366; Conservative 75; Indels 0; Gaps 0;

2 TKRPIITGVVSSIGKGIASSIAALIEDRGKVTITKLDPIYNDPGTMSPOHGEVYV 61
3 TWYIFVTGVVSSIGKGIASSIAALIEDRGKVTITKLDPIYNDPGTMSPOHGEVYV 62
62 TEDGATDLDLGHYERFLTKTKKNTFTGQYBOYVLENERGDTLGAIVYIPIITD 121
63 TEDGATDLDLGHYERFLTKTKKNTFTGQYBOYVLENERGDTLGAIVYIPIITD 122
122 IKRRVYSAEGKQVALIEVGTVGDIESHLPLETRROMGEVDRALPIHLTLVPIYS 181
123 IKRRVYSAEGKQVALIEVGTVGDIESHLPLETRROMGEVDRALPIHLTLVPIYS 182
182 AGEELTKPTQSHVKELRTTGIQDPIICSEOPIPASERKIALFTNAKAVISALDAD 241
183 AGEELTKPTQSHVKELRTTGIQDPIICSEOPIPASERKIALFTNAKAVISALDAD 242
242 TYRTRPILREOGDLDVYDQRLVPAADLSAMEKVVDGLHPDEVSIALVGRVYDHT 301
243 STYKLPOLVKSQGLDVLCTFRGIDPPEADLSMEDEVITEBANPGEVITGMVGTIELP 302
302 DAYKSLNEALIHAGIHTREKVOISYIDSETIEABGTAKLKNVDALIVPGFGEVGEKI 361
303 DAYKSLNEALIHAGIHTREKVOISYIDSETIEABGTAKLKNVDALIVPGFGEVGEKI 362
362 STVRFRERKIKITYLIGICLQMSAVIEPANNVGLGASHTEFLEPSPHRYIGLITERWDE 421
363 RAAQVAREKRVYVIGICLQMSAVIEPANNVGLGASHTEFLEPSPHRYIGLITERWDE 422
422 AGEELTVRDESDSLGTMRLGAOKCRKADSLAFOLYQKQVITERRHRYEPNNQYIKOLE 481
423 TGVNERERTSSDLGTMRLGASOLCHERKTKARELYGASTTERRHRYEPNNQYIKOLE 482
482 AAGKTFSGKSLDGRVETIELPEHPWFLACQHPHPTSTPRNGALLFGFTFAAANH 538
483 KALGVSGSLADKRVETIELPEHPWFLACQHPHPTSTPRNGALLFGFTFAAANH 539

RESULT 6

PYRG_XANAC STANDARD; PRT; 554 AA.

AC Q8PLS3; 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN PYRG OR XAC17176.
OS Xanthomonas axonopodis (pv. citrli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.

OX NCBI_Taxid=92829;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;

RX MEDLINE=22022145; PubMed=12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Furian L.R.,

RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,

RA Camarotte G., Canavan F., Cardoso J., Chamberg F., Clapina L.P.,

RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

RA Foroughieri R.F., Franco M.C., Greggio C.C., Gruber A.,

RA Kariyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Meidanis J., Wenck C.F.W., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Texeira R.I.D.,

RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,

RA Serubal J.C., Klatjima J.P.;

RA "Comparison of the genomes of two Xanthomonas pathogens with differing

RT host specificities."

RL Nature 417:459-463(2002).

CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with

CC either L-glutamine or ammonia as the source of nitrogen (By

CC similarity).

CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.

CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine

CC is the substrate. Inhibited by CTP (By similarity).

CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP, third

CC (last) step.

CC -1- SUBUNIT: Homotrimer (By similarity).

CC -1- SIMILARITY: Belongs to the CTP synthase family.

CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AE011804; AAM6583.1; -

DR HAMAP: MF_01227; -

DR InterPro: IPR000991; GATase_1.

DR InterPro: IPR004468; Pyrg.

DR Pfam: PF00117; GATase, 1.

DR TIGRFAMs: TIGR00337; Pyrg, 1.

DR PROSITE: PS00442; GATASE_TYPE_1; 1.

DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.

FT DOMAIN 1 301 AMINATOR DOMAIN
FT ACT_SITE 302 554 GUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 380 380 GATASE (BY SIMILARITY).
FT ACT_SITE 518 518 GATASE (BY SIMILARITY).
FT ACT_SITE 520 520 GATASE (BY SIMILARITY).
SQ SEQUENCE 554 AA; 61482 MW; 80DD42A9DBA23596 CRC64;

Query Match 69.3%; Score 1931; DB 1; Length 554;
Best Local Similarity 67.0%; Pred. No. 1.2e-113; Mismatches 97; Indels 4; Gaps 2;

Matches 366; Conservative 79; Indels 4; Gaps 2;

1 MTKPIITGVVSSIGKGIASSIAALIEDRGKVTITKLDPIYNDPGTMSPOHGEVYV 60
1 MTKPIITGVVSSIGKGIASSIAALIEDRGKVTITKLDPIYNDPGTMSPOHGEVYV 60
61 VTRDGAETDLDLGHYERFLTKTKKNTFTGQYBOYVLENERGDTLGAIVYIPIITD 120
61 VTRDGAETDLDLGHYERFLTKTKKNTFTGQYBOYVLENERGDTLGAIVYIPIITD 120
121 EIKRVSASAGKQVALIEVGTVGDIESHLPLETRROMGEVDRALPIHLTLVPIYK 180

```

Db 121 EIRRIDATAGPVALIEIGTVDISLPLFAIRQVTERGAEMPHLTVPIYA 180
Qy 181 SAGELKTEPGHSHVKELETIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 240
Db 181 AAGELKTEPGHSHVKELETIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 240
Qy 241 DTVYIPILLRBOGJDDIIVVQOLR--DYPAAADISAMERKVDGLTHPDEVSIAIVGKYV 298
Db 241 DTVYIPILLRBOGJDDIIVVQOLR--DYPAAADISAMERKVDGLTHPDEVSIAIVGKYV 298
Qy 239 DHTDAKSLMEALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDALIVGSGFERGVE 358
Db 239 DHTDAKSLMEALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDALIVGSGFERGVE 358
Qy 301 DHQDAKSVAAELRBRGHRQRTKVKLBAQDLBGSMSALQDIDGLIVGSGFERGVE 360
Db 301 DHQDAKSVAAELRBRGHRQRTKVKLBAQDLBGSMSALQDIDGLIVGSGFERGVE 360
Qy 359 GKISTVRARERKILTYLIGICLQMSAVIERANVVGLEGAISTEPKPSHPVIGLITW 418
Db 359 GKISTVRARERKILTYLIGICLQMSAVIERANVVGLEGAISTEPKPSHPVIGLITW 418
Qy 361 GKVOQSKPARERQKVEYFGICGMQAAVVDYAHVADLDAASTENDRQSPHPVIGLITW 420
Db 361 GKVOQSKPARERQKVEYFGICGMQAAVVDYAHVADLDAASTENDRQSPHPVIGLITW 420
Qy 419 MDAGELVTRDESDIGGTMRLGAQCRKADSLAFOLYOKDVTERRHREYEPNNQYLK 478
Db 419 MDAGELVTRDESDIGGTMRLGAQCRKADSLAFOLYOKDVTERRHREYEPNNQYLK 478
Qy 421 RTAGEVERKEDKSLDGLGTMRLGLOEORLKRGLARVYAGDVAAERHREYEPNNQYLK 480
Db 421 RTAGEVERKEDKSLDGLGTMRLGLOEORLKRGLARVYAGDVAAERHREYEPNNQYLK 480
Qy 479 QLEAAGMKPSGKSLDGLRVLVEITLPE--HPWELACQFPEFTSTPRNGHALPSGFVEAA 536
Db 479 QLEAAGMKPSGKSLDGLRVLVEITLPE--HPWELACQFPEFTSTPRNGHALPSGFVEAA 536
Qy 481 QLEDAGLVIGSKSMDDTVVEVVELPRDTHPWLACQALPEFLSTPRDGHPLFTGFVRAAR 540
Db 481 QLEDAGLVIGSKSMDDTVVEVVELPRDTHPWLACQALPEFLSTPRDGHPLFTGFVRAAR 540
Qy 537 KHKQOG 542
Db 541 EKKAGG 546

```

RESULT 7

```

ID PYRG VIBVU STANDARD; PRT; 545 AA.
AC 08DC63;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN PYRG OR VV11578.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrrio.
OX NCBI_TaxID=672;
RN RP
RA SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
either L-glutamine or ammonia as the source of nitrogen (By
similarity).
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
is the substrate; inhibited by CTP (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UTP to CTP; third
(last) step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the CTP synthase family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
CC EMBL; AE016802; AAC01002.1; -
CC HAMAP; MF_01227; -; 1.

```

```

DR InterPro; IPR000951; GATase_1.
DR InterPro; IPR004468; Pyrg.
DR Pfam; PF00117; GATase; 1.
DR TIGRPFAM; TIGR00337; Pyrg; 1.
DR PROSITE; PS00442; GATase Type I; 1.
KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.
FT DOMAIN 1 300
FT ACT_SITE 301 545
FT ACT_SITE 379 379
FT ACT_SITE 515 515
FT ACT_SITE 517 517
SO SEQUENCE 545 AA; 59872 MW; 0E71A40595F2805 CRC64;

Query Match 69.0%; Score 1925; DB 1; Length 545;
Best Local Similarity 67.0%; Pred. No. 2,8e-113;
Matches 360; Conservative 80; Mismatches 97; Indels 0; Gaps 0;

```

```

Qy 2 TKPFTIGGVVSSLSKGIASLSAILEDRLKVTITKLDPIYINVPQTMSPFQHGVEV 61
Db 3 TNYIFVGVVSSLSKGIASLSAILEDRLKVTITKLDPIYINVPQTMSPFQHGVEV 62
Qy 62 TEDGAEFDLGLHYERFLTKTKKNNFTQGVYBOYLNERKGDYAGATVQVIPHITDE 121
Db 63 TEDGAEFDLGLHYERFLTKTKKNNFTQGVYBOYLNERKGDYAGATVQVIPHITND 122
Qy 122 IKRVSASAGKQVALIENGQVGDIESLPELTITQMGVLEGRDALFHLTLVPYKS 181
Db 123 IKRVSASAGKQVALIENGQVGDIESLPELTITQMGVLEGRDALFHLTLVPYLA 182
Qy 182 AGEIKTEPTQSHVKELETIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 241
Db 183 AGEIKTEPTQSHVKELETIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 242
Qy 242 TIRIPILLRBOGJDDIIVVQOLR--DYPAAADISAMERKVDGLTHPDEVSIAIVGKYVDHT 301
Db 243 TIRIPILLRBOGJDDIIVVQOLR--DYPAAADISAMERKVDGLTHPDEVSIAIVGKYVDHT 302
Qy 302 DAVKSLMEALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDALIVGSGFERGVEGKI 361
Db 303 DAVKSLMEALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDALIVGSGFERGVEGKI 362
Qy 362 STVRPARENKIPYLGICLQMSAVIERANVVGLEGAISTEPKPSHPVIGLITWMD 421
Db 363 LAATYARENKIPYLGICLQMSAVIERANVVGLEGAISTEPKPSHPVIGLITWMD 422
Qy 422 AGEIVTRDESDIGGTMRLGAQCRKADSLAFOLYOKDVTERRHREYEPNNQYLKOLE 481
Db 423 AGEIVTRDESDIGGTMRLGAQCRKADSLAFOLYOKDVTERRHREYEPNNQYLKOLE 482
Qy 482 AAGMKPSGKSLDGLRVLVEITLPE--HPWELACQFPEFTSTPRNGHALPSGFVEAAAKH 538
Db 483 AAGMKPSGKSLDGLRVLVEITLPE--HPWELACQFPEFTSTPRNGHALPSGFVEAAAKH 539

```

RESULT 8

```

ID PYRG VIBCH STANDARD; PRT; 544 AA.
AC 09KPC4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN PYRG OR VC2448.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN RP
RA SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

```

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA McQuaen M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uitterlinden T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 CC either L-glutamine or ammonia as the source of nitrogen (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 CC is the substrate. Inhibited by CTP (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
 CC (last) step.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SIMILARITY: Belongs to the CTP synthase family.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE004315; AAF95590.1; -;
 CC FIR; E82074; E82074.
 CC TIGR; VC2448; B82074.
 DR HAMAP: MF_01227; -; 1.
 DR InterPro: IPR000991; GATase_1.
 DR Pfam: PF00117; GATase_1.
 DR PROSITE; PS00442; GATASE_TYPE_1; 1.
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
 KM Complete proteome.
 FT INIT MET 0 0 BY SIMILARITY.
 FT DOMAIN 1 299 AMINATOR DOMAIN.
 FT ACT SITE 300 544 GLUTAMINE AMIDOTRANSFERASE.
 FT ACT SITE 378 378 GATASE (BY SIMILARITY).
 FT ACT SITE 514 514 GATASE (BY SIMILARITY).
 FT ACT SITE 516 516 GATASE (BY SIMILARITY).
 SQ SEQUENCE 544 AA; 58754 MW; 3ACBEP85C35398 CRC64;
 Query Match 69.0%; Score 1924; DB 1; Length 544;
 Best Local Similarity 67.5%; Pred. No. 3.2e-113;
 Matches 362; Conservative 76; Mismatches 98; Indels 0; Gaps 0;
 Oy 2 TKRFTFGVSSLGKIAASSLAILEBGRKVTITKLDPIYINVDPGTSPFGHGVFV 61
 Db 2 TNYITFGVSSLGKIAASSLAILEBGRKVTITKLDPIYINVDPGTSPFGHGVFV 61
 Oy 62 TEDGATFDLDLGHYERPLKTKKNNFTTGQYEVGLNENKGDYLGATVYIPIITDE 121
 Db 62 TEDGATFDLDLGHYERPLKTKKNNFTTGQYEVGLNENKGDYLGATVYIPIITDE 121
 Oy 122 IRRVYVESAGDVALLIEVGTVGDIESTPELETHOMVEVGRDGLTIHTIVPYKS 181
 Db 122 IKRVYVESAGDVALLIEVGTVGDIESTPELETHOMVEVGRDGLTIHTIVPYKS 181
 Oy 122 IKRVYVESAGDVALLIEVGTVGDIESTPELETHOMVEVGRDGLTIHTIVPYKS 181
 Db 122 IKRVYVESAGDVALLIEVGTVGDIESTPELETHOMVEVGRDGLTIHTIVPYKS 181
 Oy 182 AGELEKTPQHSYKELRTIGDPIILICHSBOPIPASERKIALFTNVAEKAVISADAD 241
 Db 182 AGELEKTPQHSYKELRTIGDPIILICHSBOPIPASERKIALFTNVAEKAVISADAD 241
 Oy 242 TYVRPILLAREGLDVIYDQRLDVPADLSMENVUGLTHPDRVSIATVGVKVDHT 301
 Db 242 TYVRPILLAREGLDVIYDQRLDVPADLSMENVUGLTHPDRVSIATVGVKVDHT 301
 Oy 242 STYKLPOLRSQGLDVCARFQINAPKADLSMEVQVITTEANPGEVITGVGKTELP 301
 Db 242 STYKLPOLRSQGLDVCARFQINAPKADLSMEVQVITTEANPGEVITGVGKTELP 301
 Oy 302 DAVKSLNEALIHAGITRRKVOISYIDSETIEAGTAKKANVDAILVPGFGEVGEKTI 361
 Db 302 DAVKSLNEALIHAGITRRKVOISYIDSETIEAGTAKKANVDAILVPGFGEVGEKTI 361
 Oy 302 DAVKSLNEALIHAGITRRKVOISYIDSETIEAGTAKKANVDAILVPGFGEVGEKTI 361
 Db 302 DAVKSLNEALIHAGITRRKVOISYIDSETIEAGTAKKANVDAILVPGFGEVGEKTI 361

Oy 362 STVREARENKIPYICICIMOSAVIEFARNVYEGASSTEEFLPKSPHPIVGLITEMWDE 421
 Db 362 STVREARENKIPYICICIMOSAVIEFARNVYEGASSTEEFLPKSPHPIVGLITEMWDE 421
 Oy 422 ABEVYTRBEDSLGGTMRGAKKRLKADSLAFOLYKQDVITERRHRYEFNNQYLOKE 481
 Db 422 ABEVYTRBEDSLGGTMRGAKKRLKADSLAFOLYKQDVITERRHRYEFNNQYLOKE 481
 Oy 482 AAGMPSGKSLDGRVLEIIEHPHFLACOFHPFSTSPENGALFSGFVAAAK 537
 Db 482 AAGMPSGKSLDGRVLEIIEHPHFLACOFHPFSTSPENGALFSGFVAAAK 537
 Oy 482 KAGLAVSGLSNDKLVVYIENPAPHPVFAAOPHEFTSTPDGHLFPGFVAAAG 537
 Db 482 KAGLAVSGLSNDKLVVYIENPAPHPVFAAOPHEFTSTPDGHLFPGFVAAAG 537
 RESULT 9
 ID PYRG_PASMU STANDARD; PRT; 542 AA.
 AC 09CJG9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 GN PYRG OR PM1872.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxId=747;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Em70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of *Pasteurella multocida* Em70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 CC either L-glutamine or ammonia as the source of nitrogen (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 CC is the substrate. Inhibited by CTP (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
 CC (last) step.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SIMILARITY: Belongs to the CTP synthase family.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE006225; AAK03956.1; -;
 DR HAMAP: MF_01227; -; 1.
 DR InterPro: IPR000991; GATase_1.
 DR Pfam: PF00117; GATase_1.
 DR TIGRfam: TIGR00337; Pyrg_1.
 DR PROSITE; PS00442; GATASE_TYPE_1; 1.
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
 KM Complete proteome.
 FT DOMAIN 1 300 AMINATOR DOMAIN.
 FT ACT SITE 301 542 GLUTAMINE AMIDOTRANSFERASE.
 FT ACT SITE 379 379 GATASE (BY SIMILARITY).
 FT ACT SITE 515 515 GATASE (BY SIMILARITY).
 FT ACT SITE 517 517 GATASE (BY SIMILARITY).
 SQ SEQUENCE 542 AA; 59801 MW; CZA72913A502A612 CRC64;
 Query Match 68.8%; Score 1918; DB 1; Length 542;
 Best Local Similarity 68.4%; Pred. No. 7.6e-113;
 Matches 365; Conservative 68; Mismatches 101; Indels 0; Gaps 0;

QY 2 TKF1PITGVVSLKGIKIAASLAALIEDRGKLTITLADYINVDPMGNSPFGHGVFV 61
 DB 3 TNYIFVTGGVSSLKGIKIAASLAALIEDRGKLTITLADYINVDPMGNSPFGHGVFV 62
 QY 62 TEDGERTDLDGHERELKMTKKNFTTGOVYEQVIRNRKDYAGATOVPHITDE 121
 DB 63 TQDGEATDLDGHERELKMTKKNFTTGOVYEQVIRNRKDYAGATOVPHITDE 122
 QY 122 IKRRTVESABKDVALLIEVGTVEDIESLPLETIRQWVEAGSDRALFTMLTVPYIKS 181
 DB 123 IKSRVTDGAAGHDVALIVEGTVGDIISLPLETIRQWVEAGSDRALFTMLTVPYIKS 182
 QY 182 AGEKTKRQTSVVELRTITGTPDILIRSRQPIPASRKRILAFITNAKAVSAIDAD 241
 DB 183 AGEVTKRQTSVVELRTITGTPDILIRSRQPIPASRKRILAFITNAKAVSAIDAD 242
 QY 242 TTYRPLLRQGLDLDVVDQRLDVPADLSAMEKVVDGLTHTDEVSAIVKQYDHT 301
 DB 243 STYQIPALKXQGLDLDVVDQRLDVPADLSAMEKVVDGLTHTDEVSAIVKQYDHT 302
 QY 302 DAYVSNMALIHAGITHHKKVQIYISIDETIABSTAKLVNDALVPGGGERGVBSKI 361
 DB 303 DAYVSNMALIHAGITHHKKVQIYISIDETIABSTAKLVNDALVPGGGERGVBSKI 362
 QY 362 STVRPARENKIPLYGICGMSAVIEFARNVVGEGASTEFPLSPHPVIGITENWDE 421
 DB 363 LTAKARKNEIPLYGICGMSAVIEFARNVVGEGASTEFPLSPHPVIGITENWDE 422
 QY 422 AGEVTRDESDLDGGMELGAQCKLADSLAPOLYQDVITERRHRYEFNNQLYQLE 481
 DB 423 EGNIEETRTDADLDGGMELGAQCKLADSLAPOLYQDVITERRHRYEFNNQLYQLE 482
 QY 482 AAGMKESKSLDGRVETIELPEHPMFLACQHPPEFTSPNGHAFSGFEVA 535
 DB 483 KAGLVATSLADKCLVELLEVPNHPFVACQHPPEFTSPNGHAFSGFEVA 536

RESULT 10
 ID PYRG ECOLI STANDARD; PRT: 544 AA.
 AC P08358; 01-AUG-1988 (Rel. 08, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DF CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 GN PYRG OR B2780 OR C3345 OR Z4095 OR ECS3640 OR SF2795.
 OS Escherichia coli
 OS Escherichia coli O6,
 OS Escherichia coli O157:H7, and
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562, 217992, 83334, 623;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli;
 RX MEDLINE=86168304; PubMed=3514618;
 RA Weng M., Makaroff C.A., Zaikin H.,
 RT "Nucleotide sequence of Escherichia coli pyrg encoding CTP
 RT synthetase.";
 RL J. Biol. Chem. 261:5568-5574(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]

RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Zhou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.,
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Pofal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., DiMallanca E.T., Potamowski K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.,
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Kuida S., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.,
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL D.N.A. Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 1-12.
 RC SPECIES=E.coli; STRAIN=K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.U., Robison K., Church G.M.,
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=2272406; PubMed=12384559;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu M., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Hu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.,
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [8]
 RP MOTAGENESIS OF VAL-346; GLY-350 AND GLY-351.
 RC SPECIES=E.coli;
 RX MEDLINE=87250264; PubMed=3298209;
 RA Weng M., Zaikin H.,
 RT "Structural role for a conserved region in the CTP synthetase
 RT glutamine amide transfer domain.";
 RL J. Bacteriol. 169:3023-3028(1987).
 RN [9]
 RP ENZYME REGULATORS.
 RC SPECIES=E.coli;
 RX MEDLINE=9322179; PubMed=8385490;
 RA Robertson J.G., Villafraña J.U.,
 RT "Characterization of metal ion activation and inhibition of CTP
 RT synthetase.";
 RL Biochemistry 32:3769-3777(1993).
 RN [10]
 RP ROLE OF GTP, AND MOTAGENESIS OF CYS-378.
 RC SPECIES=E.coli;
 RX MEDLINE=21234865; PubMed=11336655;

Query Match	Best Local Similarity	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231	Length	DB 232	Score	DB 233	Length	DB 234	Score	DB 235	Length	DB 236	Score	DB 237	Length	DB 238	Score	DB 239	Length	DB 240	Score	DB 241	Length	DB 242	Score	DB 243	Length	DB 244	Score	DB 245	Length	DB 246	Score	DB 247	Length	DB 248	Score	DB 249	Length	DB 250	Score	DB 251	Length	DB 252	Score	DB 253	Length	DB 254	Score	DB 255	Length	DB 256	Score	DB 257	Length	DB 258	Score	DB 259	Length	DB 260	Score	DB 261
-------------	-----------------------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------

QY	2	TKKIFIFIGVAVSSLGGIGIAASSLAAILBERGAKVITTYLIDYINWDPCMSPPGHGVFV	6
QY <td>2</td> <td>TKKIFIFIGVAVSSLGGIGIAASSLAAILBERGAKVITTYLIDYINWDPCMSPPGHGVFV</td> <td>6</td>	2	TKKIFIFIGVAVSSLGGIGIAASSLAAILBERGAKVITTYLIDYINWDPCMSPPGHGVFV	6
Db	2	TNTVFTVIGGVAVSSLGGIGIAASSLAAILBERGAKVITTYLIDYINWDPCMSPPGHGVFV	61
QY	62	TEQDCAETDLDLGHYERFELTKTNTKKNFTTGGVQEQVLRNKKQDYGATVQVTPHITDB	121
QY	62	TEQDCAETDLDLGHYERFELTKTNTKKNFTTGGVQEQVLRNKKQDYGATVQVTPHITDB	121
QY	122	IKQRYVESASGRQVALIEVGVTVGDIESTLPEFLRTIRQMGVELGRDRLPHILTLVPIYKS	181
Db	122	IKQRYVESASGRQVALIEVGVTVGDIESTLPEFLRTIRQMGVELGRDRLPHILTLVPIYKS	181
QY	182	AGELTKTPQTHSVKELRTIGIOPDILIKRSEQPIPASERRKIALFTVAEKAVISAIDAD	241
QY	182	AGELTKTPQTHSVKELRTIGIOPDILIKRSEQPIPASERRKIALFTVAEKAVISAIDAD	241
QY	242	TTIRIFELLREQGLDLDLVVDQRLRVPAADLSANEKVYDGLTHTPTDERSLAIYKQYDHT	301
Db	242	TTIRIFELLREQGLDLDLVVDQRLRVPAADLSANEKVYDGLTHTPTDERSLAIYKQYDHT	301
QY	302	DAYKSLNEALIHAGITRHKVQISYIDSEITLAEGETAKLKVADAILVPGQSGEEVGEKGI	361
Db	302	DAYKSLNEALIHAGITRHKVQISYIDSEITLAEGETAKLKVADAILVPGQSGEEVGEKGI	361
QY	362	STYRPAPEBKIPYLIGICLQMGQAVTEPARNAVGLBGAHSTFELKSPHPVIGLITTWMD	421
Db	362	STYRPAPEBKIPYLIGICLQMGQAVTEPARNAVGLBGAHSTFELKSPHPVIGLITTWMD	421
QY	422	AGSLAVTRDESDLGGMTMRGLAQCKLXADSLAPQLYQXQDVITTEHRHRYEPNNQYLLQLE	481
Db	422	AGSLAVTRDESDLGGMTMRGLAQCKLXADSLAPQLYQXQDVITTEHRHRYEPNNQYLLQLE	481
QY	482	AAGAKTSKSGSLDGRVLEITELPEHFVFLACOFHEPFTSTPENGHALSGFEVAAK	537
Db	482	AAGAKTSKSGSLDGRVLEITELPEHFVFLACOFHEPFTSTPENGHALSGFEVAAK	537
RESULT 11			
PGYR_YERPE	STANDARD;	PRT;	544 AA.
AC	Q02BN1;		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).		
GN	PGYR OR YPOJ377 OR Y0813.		
OS	Yersinia pestis.		
OC	Bacteriia; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Yersinia.		
OX	NCBI_TaxID=632;		
RA	SEQUENCE FROM N.A.		
RP	STRAIN=CO-92 / Biovar Orientalis;		
RC	MEDLINE=21470413; PubMed=11586360;		
RA	Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,		
RA	Prentice M.B., Sebahnia M., James K.D., Churcher C., Mungall K.L.,		
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,		
RA	Chillingworth T., Cronin A., Davies K.M., Davis P., Dougan A.V.,		
RA	Feltwell T., Hamlin N., Holtroyd S., Jagers K., Karlyshev A.V.,		
RA	Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,		
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,		
RT	"Genome sequence of Yersinia pestis, the causative agent of plague.";		
RL	Nature 413:523-527[2001].		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KIMS / Biovar Mediaevalis;		
RA	MEDLINE=22137863; PubMed=12142430;		
RA	Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Lies P.,		
RA	Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,		
RA	Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,		
RA	Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,		
RT	Perry R.D.;		
RT	"Genome sequence of Yersinia pestis KIM.";		


```

RL J. Bacteriol. 184:4601-4611 (2002).
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
CC is the substrate. Inhibited by CTP (by similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
CC (last) step.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SIMILARITY: Belongs to the CTP synthase family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ14156; CAC92607.1; -.
DR EMBL: AB013684; AAM84400.1; -.
DR PIR: AC0410; AC0410.
DR HAMAP: MF_01227; -; 1.
DR InterPro: IPR000991; GATase_1.
DR Pfam: PF00117; GATase_1.
DR PROSITE: PS00442; GATASE_TYPE_1.
DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.
FT INIT MET 0 BY SIMILARITY.
FT DOMAIN 1 299 AMINATOR DOMAIN.
FT DOMAIN 300 544 GLUTAMINE AMIDOTRANSFERASE.
FT ACT SITE 378 378 GATASE (BY SIMILARITY).
FT ACT SITE 514 514 GATASE (BY SIMILARITY).
FT ACT SITE 516 516 GATASE (BY SIMILARITY).
SQ SEQUENCE 544 AA; 60232 MW; A5B8A2583C5C001 CRC64;

Query Match 68.5%; Score 1909; DB 1; Length 544;
Best Local Similarity 66.2%; Pred. No. 2, 8e-112;
Matches 356; Conservative 80; Mismatches 102; Indels 0; Gaps 0;

QY 2 TKFIPTGGVSSLGKIAASLAALIEDRGKVTITKLDPIYNDPGTMSFGGEVTV 61
DB 2 TNYIPTGVGVSSLGKIAASLAALIEDRGKVTITKLDPIYNDPGTMSFGGEVTV 61
QY 62 TEDGAEITDLDGHEERELKTTMKNNFTTGVYBQVLENERGDIAGTQVYIHITD 121
DB 62 TEDGAEITDLDGHEERELKTTMKNNFTTGVYBQVLENERGDIAGTQVYIHITD 121
QY 122 IKRRVESAEGKQVALIEVGTVGDIESTLPELTTIRMGVELEGDRALFIHLTVPIKRS 181
DB 122 IKRRVESAEGKQVALIEVGTVGDIESTLPELTTIRMGVELEGDRALFIHLTVPIKRS 181
QY 122 IKRRITIEGGGHVVAIVEIGTVGDIESTLPELAIKQMAVDQREHTLVMLTLVPIYLA 181
DB 122 IKRRITIEGGGHVVAIVEIGTVGDIESTLPELAIKQMAVDQREHTLVMLTLVPIYLA 181
QY 182 AGSLTKRPQSHVKEKRTIGIOPDILICSEQPIASERBKALFTNVAKVISALMD 241
DB 182 AGSLTKRPQSHVKEKRTIGIOPDILICSEQPIASERBKALFTNVAKVISALMD 241
QY 182 AGSVAKKTPQSHVKEKRTIGIOPDILICSDRAVPNERAKTALFQNVESKAVISLKD 241
DB 182 AGSVAKKTPQSHVKEKRTIGIOPDILICSDRAVPNERAKTALFQNVESKAVISLKD 241
QY 242 TIYRIPLLEEQGLDLDVVDQRLDVPADLSAMERKVDGLTPTDEVALIVGKVDHT 301
DB 242 TIYRIPLLEEQGLDLDVVDQRLDVPADLSAMERKVDGLTPTDEVALIVGKVDHT 301
QY 242 SIYKIGLAKSGGLDLYICRFSLTCPENALAMEGVLEESNGEVIITGIGKVELP 301
DB 242 SIYKIGLAKSGGLDLYICRFSLTCPENALAMEGVLEESNGEVIITGIGKVELP 301
QY 302 DAYKSLNEALIHGITHRKVQISYTDSETIAGGAKIKNDALLVPGSGRGRVYKCT 361
DB 302 DAYKSLNEALIHGITHRKVQISYTDSETIAGGAKIKNDALLVPGSGRGRVYKCT 361
QY 302 DAYKSVIEALKKQGLKNRLETVNIKLDISQDVETRGEMKEKLDALIPGFGVGRVEKY 361
DB 302 DAYKSVIEALKKQGLKNRLETVNIKLDISQDVETRGEMKEKLDALIPGFGVGRVEKY 361
QY 362 STVRFARANKIPLYGICLGQSAVIEFARVVGLEGAHSTEFLPKSPHPIVIGITWMD 421
DB 362 STVRFARANKIPLYGICLGQSAVIEFARVVGLEGAHSTEFLPKSPHPIVIGITWMD 421
QY 362 IAAVYAREHNIPTLGICLGQVALMEFARVVGMEKARNTSTETPDDKTYVVALITWRE 421
DB 362 IAAVYAREHNIPTLGICLGQVALMEFARVVGMEKARNTSTETPDDKTYVVALITWRE 421
QY 422 AGEIIVREDDSDIGTMRICAKCKADSLAFQLYQKDVITERRHRYEPNNQYIKOLE 481
DB 422 AGEIIVREDDSDIGTMRICAKCKADSLAFQLYQKDVITERRHRYEPNNQYIKOLE 481
QY 422 DGNVEIRTERSDIGTMRVGGQCHLTESLVRQMGEPITVERHRRRYEVNNMLKOLE 481
DB 422 DGNVEIRTERSDIGTMRVGGQCHLTESLVRQMGEPITVERHRRRYEVNNMLKOLE 481

```

```

QY 482 AAGKFKSGKSLDGLVEIIBPEWPFACQFHEFTSTPRNGHALESQVFAAKKH 539
DB 482 AAGIRVAGRSADNKLVEIIEIPDPWPFACQFHEFTSTPRDGHPLFAGFYKAGDYQ 539

RESULT 12
PRG: HAEIN STANDARD; PRT; 545 AA.
ID: P44341.
AC: P44341.
DT: 01-NOV-1995 (Rel. 32, Created)
DT: 01-NOV-1995 (Rel. 32, Last sequence update)
DT: 28-FEB-2003 (Rel. 41, Last annotation update)
DE: CTP synthase (EC 6.3.4.2) (UTP-ammonia ligase) (CTP synthetase).
GN: PYRQ OR H1077.
OS: Haemophilus influenzae.
OC: Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC: Pasteurellaceae; Haemophilus.
OX: NCBI_TaxID=727;
RN: [1]
RP: SEQUENCE FROM N.A.
RC: STRAIN=Rd / KW20 / ATCC 51907;
RX: MEDLINE=95350630; PubMed=7542800;
RA: Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA: Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA: McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA: Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA: Weisman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,
RA: Utecherbick T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA: Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA: Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA: Venter J.C.;
FT: "Whole-genome random sequencing and assembly of Haemophilus influenzae
FT Rd."
RL: Science 269:496-512(1995).
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
CC is the substrate. Inhibited by CTP (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
CC (last) step.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SIMILARITY: Belongs to the CTP synthase family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32788; AAC22733.1; -.
DR PIR: F64181; F64181.
DR TIGR: H1077; -.
DR HAMAP: MF_01227; -; 1.
DR InterPro: IPR000991; GATase_1.
DR InterPro: IPR004468; PYRQ.
DR Pfam: PF00117; GATase_1.
DR TIGRfam: TIGR00337; PYRQ_1.
DR PROSITE: PS00442; GATASE_TYPE_1.
KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KW Complete proteome.
FT DOMAIN 1 300 AMINATOR DOMAIN.
FT DOMAIN 301 545 GLUTAMINE AMIDOTRANSFERASE.
FT ACT SITE 379 379 GATASE (BY SIMILARITY).
FT ACT SITE 515 515 GATASE (BY SIMILARITY).
FT ACT SITE 517 517 GATASE (BY SIMILARITY).
SQ SEQUENCE 545 AA; 60263 MW; 9DBBA18AF811E3CA6 CRC64;

```

Query Match 68.5%; Score 1908.5; DB 1; Length 545;
 Best Local Similarity 67.5%; Pred. No. 3e-112;
 Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

2 TXRFITFGVSSISGKGIASSLAALIEDRGKATITTKIDPIYINVDGMSPPHGEYV 61
 3 TNYIFETGVSSISGKGIASSLAALIEDRGKATITTKIDPIYINVDGMSPPHGEYV 62
 62 TEDGATDLDLGHYERFLKTMKKNFTTGOVEYOLNERNKGDYLGATVQVPHITD 121
 63 TCGAETDLDLGHYERFIRTKTKNNFTTKYSEVLRNRRRGVLTGATVQVPHITD 122
 122 IKRRVYSAEGKVALIEVGTGVDIESLPLETITOMGVELGRDALFIHLTPVYIS 181
 123 IKORVIAQSHDVIVVEGVTVGDIESTPFEALSQLAVQVGERHTLPMHLTPVYPT 182
 182 AGELKTPHOSVKEKRTIGIQPDILICSEOPIPASERKIALFTVAEKAVISADAD 241
 183 AGEVKTPTQHSVKEKRTIGIQPDILICSDNMPNERAKIALFCNVAERAVISADAD 242
 242 TIYRPLILREOGDLVDVQRLDVPADLSAMEKVDGLTHPTDEVSIALVGYVHT 301
 243 SIYQIFALKKSCQGLDPEVCEFRITCPEADLTMEGVLYKQANPVEVTIGWVGYTELP 302
 302 DAKSINELALIHAGIHRHVKVOISYIDSEITAEGRNAXIKNDALITVPGSPGREGKI 361
 303 DAKSVNALKKAGLITNRLSVNITKIDSOVETKGVYVILKIDGILVPGSPGREGKI 362
 362 STYVFARENKIPYLGICLQVOSAVIEFPAENVGLRGAHSTEEPSPKSPHPIGLITEMWD 421
 363 RPAQVARENKIPYLGICLQVQALIEYASNVAGLITFANSSEPKDCEGVVALITEMQDA 422
 422 AGEIIVRDESDSLGCTMRIGAKCRKADSLAQYQKOVITERRHRRHPNNQYKOL 481
 423 EGVTEVTRDESIDGCTMRIGAKCRKADSLAQYQKOVITERRHRRHPNNQYKOL 482
 482 AAGMKFSKSLDGLVEIIELEPHFPLACQFHEFTSTPRNHALSGFEVAA-AKHK 539
 483 KGLKATYGLSADKGLVEIIEVPHFPAVACQFHEFTSTPRNHALSGFEVAA-AKHK 541

RESULT 13
 PYRG_XYLF STANDARD; PRT; 554 AA.

AC 09PDI,
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 GN PYRG OR XF1288.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 NCBI_TaxID=2371;

SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Alvarado A.J.G., Reinach F.C., Arruda P., Abreu F.A., Apiceno M.,
 RA Alvarado R.J., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britona M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colatto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Perro J.A.,
 RA Fagari J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
 RA Ho P.L., Honeisel J.D., Jungueira M.L., Kemper B.L., Katsjima J.P.,
 RA Krieger J.B., Kuzmae B.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Marques M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins S.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Mitraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhami A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Pais A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pasquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Sannelli R.V., Sawasaki H.F.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva M.A.Jr.,
 RA da Silveira J.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldman J., Zetser J.C.,
 RT The genome sequence of the plant pathogen *Xylella fastidiosa*.
 RL Nature 406:151-159(2000).
 CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 CC either L-glutamine or ammonia as the source of nitrogen (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 CC is the substrate. Inhibited by CTP (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
 CC (last) step.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SIMILARITY: Belongs to the CTP synthase family.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AE003962; AAF64097.1; -
 DR PIR; C82700; C82700.
 DR HAMAP; MF_01227; -; 1.
 DR InterPro; IPR000991; Glutase_1.
 DR InterPro; IPR004468; PYRG.
 DR Pfam; PF00117; GATase; 1.
 DR TIGRFAMs; TIGR00337; Pyrg; 1.
 DR PROSITE; PS00442; GATase_type_1; 1.
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
 KW Complete proteome.
 KW DOMAIN 1 301 AMINATOR DOMAIN.
 FT DOMAIN 302 554 GLUTAMINE AMIDOTRANSFERASE.
 FT ACT_SITE 380 380 GATASE (BY SIMILARITY).
 FT ACT_SITE 518 518 GATASE (BY SIMILARITY).
 FT ACT_SITE 520 520 GATASE (BY SIMILARITY).
 SQ SEQUENCE 554 AA; 61558 MW; 5E8769917043811 CRC64;

Query Match 68.3%; Score 1905; DB 1; Length 554;
 Best Local Similarity 66.1%; Pred. No. 5.1e-112;
 Matches 361; Conservative 79; Mismatches 102; Indels 4; Gaps 2;

1 MTKFIFITGGVSSISGKGIASSLAALIEDRGKATITTKIDPIYINVDGMSPPHGEYV 60
 1 MTKFIFITGGVSSISGKGIASSLAALIEDRGKATITTKIDPIYINVDGMSPPHGEYV 60
 61 VTDGATDLDLGHYERFLKTMKKNFTTGOVEYOLNERNKGDYLGATVQVPHITD 120
 61 VTDGATDLDLGHYERFLKTMKKNFTTGOVEYOLNERNKGDYLGATVQVPHITD 120
 121 IKRRVYSAEGKVALIEVGTGVDIESLPLETITOMGVELGRDALFIHLTPVYIK 180
 121 IKRRVYSAEGKVALIEVGTGVDIESLPLETITOMGVELGRDALFIHLTPVYIK 180
 121 EIRRCIDATASFDVALVEIGVTVGDIESTPFEALSQLAVQVGERHTLPMHLTPVYPT 180
 181 SAGEKTPHOSVKEKRTIGIQPDILICSEOPIPASERKIALFTVAEKAVISADAD 240
 181 SAGEKTPHOSVKEKRTIGIQPDILICSEOPIPASERKIALFTVAEKAVISADAD 240
 181 AAGBKTPHOSVKEKRTIGIQPDILICSDNMPNERAKIALFCNVAERAVISADAD 240
 241 DTIYRPLILREOGDLVDVQRLDVPADLSAMEKVDGLTHPTDEVSIALVGYVHT 298
 241 DTIYRPLILREOGDLVDVQRLDVPADLSAMEKVDGLTHPTDEVSIALVGYVHT 298
 241 DVLYGMPLELRQGLDVIIVDQFKLSGTASLADLSEMDVDVDAIKHPIDEVTIAVVGKIV 300

QY 299 DHTDAKSLNEALIHAGHTRHKVOISYDSETEAEATKLNDAIIVPGSGEGRVE 358
 DB 301 DYQADAKYSVGEALDKHOGKFORTRKVMKWEADLESDMGALDKIDGLIVPGSGFDGRPE 360
 QY 359 GKISTVPRERENKIPYIGICLQKQSAVIEFARNVGLGASHTEFLPKSPHVVIGLTEM 418
 DB 361 GKVALSRVAREQVRVPIGICQKQAAVVOYARHAGLSEANSTENDROSPPHVALITEM 420
 QY 419 MDEAGELVTRDESDSDIGTMRIGAOKRKLKADSLAFOLYQKDVITERHRRHYEPNNQYIK 478
 DB 421 RTTTCGEVERDEKSDIGTMRIGLQEQRLKAGTIVVEIYGRDVGERRHRYEPNNRYRT 480
 QY 479 QLEAAKTKSGKSLDRLVETIELPR--HPWFLACQFHEPFTSTRNHGLPSGVEAA 536
 DB 481 QLEDAELVTAASKMDTLVEMTELEPRNMPWFLACQAFPEFLSTPDGHPLTGFVKAR 540
 QY 537 KHKTOG 542
 DB 541 ARKAGG 546

RESULT 14

PRG_NEIMA STANDARD; PRT; 544 AA.

ID PYRG_NEIMA
 AC 09UTL;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 GN PYRG OR NMB154.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 NCBI_TaxID=65689;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
 Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,
 Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrell B.G.;
 "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."
 RT Nature 404:502-506 (2000).
 RL -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 either L-glutamine or ammonia as the source of nitrogen (By
 similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 is the substrate. Inhibited by CTP (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
 (last) step.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SIMILARITY: Belongs to the CTP synthase family.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL collaboration
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isd-sib.ch/announce/>
 or send an email to license@sib-sb.ch).
 CC EMBL: AL162757; CAB84970.1;
 CC PIR: F81798; P81798.
 CC HAMAP: MF_01227; 1.
 CC InterPro: IPR000991; GATase_1.
 CC InterPro: IPR004468; Pyrg.

DR Pfam: PF00117; GATase; 1.
 DR TIGRfam: TIGR00337; Pyrg; 1.
 DR Prosite: PS00442; GATase type 1; 1.
 KW Pyrimidine biosynthesis; ligase; Glutamine amidotransferase;
 KM Complete proteome.
 FT DOMAIN 1 299 AMINATOR DOMAIN.
 FT ACT_SITE 300 544 GLUTAMINE AMIDOTRANSFERASE.
 FT ACT_SITE 380 380 GATASE (BY SIMILARITY).
 FT ACT_SITE 517 517 GATASE (BY SIMILARITY).
 FT ACT_SITE 519 519 GATASE (BY SIMILARITY).
 SQ SEQUENCE 544 AA; 59916 MW; 99376E7C6557CAE2 CRC64;

Query Match 67.9%; Score 1892.5; DB 1; Length 544;
 Best Local Similarity 66.1%; Pred. No. 3e-111;
 Matches 358; Conservative 80; Mismatches 10; Indels 3; Gaps 2;

QY 1 MTKRIFINGVAVSSILGKGIASLAAILDRLGLRTYTKLDPIYINVDPGTSPFGHEVF 60
 DB 1 MTKRIFVGVAVSSILGKGIASLAAILDRLGLRTYTKLDPIYINVDPGTSPFGHEVF 60
 QY 61 VTDEGAEFTDLDGHYERFLKTTTKKNNFTGQYVEQVLRNRRGDTAGATVOYIPIHTD 120
 DB 61 VTDEGAEFTDLDGHYERFLKTTTKKNNFTGQYVEQVLRNRRGDTAGATVOYIPIHTD 120
 QY 121 EIKRVSASAEKQVALIEVGVGDIESTPLEFTIRQMGVGLGRDALFTHTLVPIYK 180
 DB 121 EIKRVSASAEKQVALIEVGVGDIESTPLEFTIRQMGVGLGRDALFTHTLVPIYK 180
 QY 121 EIKRVSASAEKQVALIEVGVGDIESTPLEFTIRQMGVGLGRDALFTHTLVPIYK 180
 DB 121 EIKRVSASAEKQVALIEVGVGDIESTPLEFTIRQMGVGLGRDALFTHTLVPIYK 180
 QY 181 SAGBLKTPYSHVYELRTIGIOPDILICSEQPIPSERKIALFTYNAKAVISALDA 240
 DB 181 SAGBLKTPYSHVYELRTIGIOPDILICSEQPIPSERKIALFTYNAKAVISALDA 240
 DB 181 AAGRIKXKPTQHTVYKEMSLIGLQPDILICRMDRTMPADERRKIALFCVEERAVIGSDV 240
 QY 241 DTIRIFPLARBOGLDLDVQDLADVPADLSAMEKVDLTHPTDEVSIAIVGYVDH 300
 DB 241 DTIRIFPLARBOGLDLDVQDLADVPADLSAMEKVDLTHPTDEVSIAIVGYVDH 300
 DB 241 DSIYCEPMLHDQSDIDNITITQLOANVOQADLTMKKIVAHIQPKTVKIAMGVYDL 300
 QY 301 TDVAKSLNEALIHAGHTRHKVOISYDSETEAE--TAKLNDAIIVPGSGEGRVE 358
 DB 301 TDVAKSLNEALIHAGHTRHKVOISYDSETEAE--TAKLNDAIIVPGSGEGRVE 358
 DB 301 TBSYKSLLEALKHAGVTEITVQITTPVDSIEIRKNGDVSLKMDAILVPGSGSRVE 360
 QY 359 GKISTVPRERENKIPYIGICLQKQSAVIEFARNVGLGASHTEFLPKSPHVVIGLTEM 418
 DB 359 GKISTVPRERENKIPYIGICLQKQSAVIEFARNVGLGASHTEFLPKSPHVVIGLTEM 418
 DB 361 GKVALSRVAREQVRVPIGICQKQAAVVOYARHAGLSEANSTENDROSPPHVALITEM 420
 QY 419 MDEAGELVTRDESDSDIGTMRIGAOKRKLKADSLAFOLYQKDVITERHRRHYEPNNQYIK 478
 DB 419 MDEAGELVTRDESDSDIGTMRIGAOKRKLKADSLAFOLYQKDVITERHRRHYEPNNQYIK 478
 DB 421 QTADGVSATREDSEADIGTMRIGLQEQRLKAGTIVVEIYGRDVGERRHRYEPNNRYRT 480
 QY 479 QLEAAKTKSGKSLD--GRVETIELPRNMPWFLACQFHEPFTSTRNHGLPSGVEAA 537
 DB 479 QLEAAKTKSGKSLD--GRVETIELPRNMPWFLACQFHEPFTSTRNHGLPSGVEAA 537
 DB 481 TLEQAGLVIGVSGARELVETIELPRNMPWFLACQFHEPFTSTRNHGLPSGVEAA 540
 QY 538 HK 539
 DB 541 NK 542

RESULT 15

PRG_NEIMA STANDARD; PRT; 544 AA.

ID PYRG_NEIMA
 AC 09UTL;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
 GN PYRG OR NMB154.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 NCBI_TaxID=491;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;

RX MEDLINE=20175755; PubMed=10710307;
 RA Terebin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Ghim M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scariato V., Masigant V., Pizzo M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappaport R., Venter J.C.,
 RA "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 MC58.";
 RL Science 287:1809-1815 (2000).
 CC - FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
 either L-glutamine or ammonia as the source of nitrogen [by
 similarity].
 CC - CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
 CC - ENZYME REGULATION: Allosterically activated by GTP, when glutamine
 is the substrate. Inhibited by CTP (by similarity).
 CC - PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
 (last) step.
 CC - SUBUNIT: Homotetramer (by similarity).
 CC - SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: AE002505; AAF41908.1; -
 DR PIR: F81070; F81070.
 DR TIGR: NMB1554; -
 DR HAMAP: MF_01227; -; 1.
 DR InterPro: IPR000991; GATase_1.
 DR InterPro: IPR004468; Pyrg.
 DR Pfam: PF00117; GATase; 1.
 DR TIGRFAMs: TIGR00337; Pyrg; 1.
 DR PROSITE: PS00442; GATASE TYPE 1; 1.
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
 KM Complete proteome.
 FT DOMAIN 1 299 AMINATOR DOMAIN.
 FT DOMAIN 300 544 GLUTAMINE AMIDOTRANSFERASE.
 FT ACT SITE 380 380 GATASE (BY SIMILARITY).
 FT ACT SITE 517 517 GATASE (BY SIMILARITY).
 FT ACT SITE 519 519 GATASE (BY SIMILARITY).
 SQ SEQUENCE 544 AA; 59923 MW; E3C67621DF52B69C CRC64;
 Query Match 67.7%; Score 1888.5; DB 1; Length 544;
 Best Local Similarity 66.2%; Pred. No. 5.4e-111;
 Matches 359; Conservative 77; Mismatches 103; Indels 3; Gaps 2;
 QY 1 MKKFFINGGVSSSGKGIASSLAIIEDRGKVTITKLDPIYNDPGTMSPROHGEVP 60
 DB 1 MKKIFVTVGVSSSGKGIASSLAIIIESRGLNVTMLKIDPIYNDPGTMSPROHGEVP 60
 QY 61 VTBDGAETDLDLGHYERFLTKTMTKKNFTTQVYEQVLAENERKDYAGATVQVLPHTD 126
 DB 61 VTBDGAETDLDLGHYERFLTKTMTKKNFTTQVYEQVLAENERKDYAGATVQVLPHTD 126
 QY 121 EIKRRVVSAGKGVALLIEVGTVGDIISLFLFTITROMGVELGRDRAFLHLLVPIYK 180
 DB 121 EIKRRVVSAGKGVALLIEVGTVGDIISLFLFTITROMGVELGRDRAFLHLLVPIYK 180
 QY 121 EIKRRVVSAGKGVALLIEVGTVGDIISLFLFTITROMGVELGRDRAFLHLLVPIYK 180
 DB 121 EIKRRVVSAGKGVALLIEVGTVGDIISLFLFTITROMGVELGRDRAFLHLLVPIYK 180
 QY 181 SAGELKTPPTQSHVYELTITIGIPDILICRSRQIPASERRKIALFTNVAEKAVISIDA 240
 DB 181 SAGELKTPPTQSHVYELTITIGIPDILICRSRQIPASERRKIALFTNVAEKAVISIDA 240
 QY 181 AAGIKTKPTQHTYKEMSLIGLQPDILICRMDRTPADERRKIALFCNVERAIVGSYDV 240
 DB 181 AAGIKTKPTQHTYKEMSLIGLQPDILICRMDRTPADERRKIALFCNVERAIVGSYDV 240
 QY 241 DTIYRIPLLLREGLDILVDQLRDVPAADLSAMEKYVDGLTEPTDEVSAIVGKTVDH 300
 DB 241 DTIYRIPLLLREGLDILVDQLRDVPAADLSAMEKYVDGLTEPTDEVSAIVGKTVDH 300
 QY 241 DSIYCEPMELHDQGDINITEQLQANVOQADLTAMKKIVHAIQNPKHIVKIAMVGKYVDL 300
 DB 241 DSIYCEPMELHDQGDINITEQLQANVOQADLTAMKKIVHAIQNPKHIVKIAMVGKYVDL 300

QY 301 TDAYKSINEALIHAGHTPRKRVQISYDSETEAE--GTAKLKNVDAILVPGFGSGRGVE 358
 DB 301 TDSYKSLIEALKHAGHTETDVOITFPDSENIENKPKDVSMLKMDAILVPGFGSGRGVE 360
 QY 359 GKISTVAPARENKIPYIGICIGMGSAYIEFARNVVGAGASTFELKSPHPVIGLITW 418
 DB 361 GKIAVRYARENVPYIGICIGMQLALIEYARDVAGLKGANSTFEDLKAAFPVALIDW 420
 QY 419 MDEAGELVTDESDLDGVTWLAGAKCRKLNDSAPQLYOKDVITERRRRYEFNNQYLX 478
 DB 421 QTADGSVEYTRDESDLDGVTWLAGAQLVETLKXGSAAATYSGHIREHRRREYVNNVVP 480
 QY 479 QLEAAKGFSGSKLD-GRUVIITLPEHPVFLAQCFPEPTSTFRNGHALFSGVPEAAK 537
 DB 481 TLEQAGLVIGVSAGRRRLVETIELPNHPWFACQFPEFTSNPKGHPLETAIVKAAIN 540
 QY 538 HK 539
 DB 541 NK 542

Search completed: January 29, 2004, 15:50:42
 Job time : 11.244 secs


```

QY 182 AGELTKPTQSHSVKELRTIGIOPDILICRSEOPIPASERKIALFTYVAKVISAIDND 241
DB 183 AGEVTKPTQSHSVKELRTIGIOPDILICRSDRIPANERKIALFTYVAKVISAIDND 242
QY 242 TYIRPILRLRBOGLDLDVVDQLRLDVPADLSAMEKYVDGLTHPTDEVSIALVGYVDHT 301
DB 243 STYKIPALRSQGLDELIVKFRSLTCREADLSSEWENVIYOGANRNGEVIAMVKIYELP 302
QY 302 DAVKSYNALIHAGIHRHKVQISYIDSETEAEAGTAKVNDALVPGGGEVGECKI 361
DB 303 DAVKSYNALIHAGIHRHKVQISYIDSETEAEAGTAKVNDALVPGGGEVGECKI 362
QY 362 STYVARENKIPYIGICIGMOSAVIEFARNVGLGASHTEFLPKSPHPVIGLITEMWDE 421
DB 363 LAAYVARENKIPYIGICIGMOSAVIEFARNVGLGASHTEFLPKSPHPVIGLITEMWDE 422
QY 422 AGEVLTRDSDLDGCTMELGAKQCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKOLE 481
DB 423 EGVNEORREHSDLDGCTMELGAKQCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKOLE 482
QY 482 AAGMKSQKSLDGRVLEIIELEPHPMFLACQFHPBETSTPRNGHALFSGFEAAAKH 538
DB 483 QAGLVPSGLSDRKLVEMLIELEPHPMFLACQFHPBETSTPRNGHALFSGFEAAAKH 539

```

RESULT 2

```

Q8EB09 PRELIMINARY; PRT; 546 AA.
ID Q8EB09
AC Q8EB09;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE CTP synthase.
OS PYRG OR SO3441.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=2297686; PubMed=1236813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
RA Meyer T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprimis M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uitterlinden T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015780; AAN56438.1; -.
DR TIGR; SO3441; -.
KM Complete proteome.
SQ SEQUENCE 546 AA; 60148 MW; 4079A1DB933CF39E7 CRC64;

```

Query Match 69.0%; Score 1925; DB 16; Length 546;
 Best Local Similarity 68.2%; Pred. No. 2e-125;
 Matches 367; Conservative 69; Mismatches 102; Indels 0; Gaps 0;

```

QY 2 TKPIFTGVSSLGKIAASSLAALIEDRGKVTITTKLDPIYNDPGTMSFGQGEVYV 61
DB 3 TKPIFTGVSSLGKIAASSLAALIEDRGKVTITTKLDPIYNDPGTMSFGQGEVYV 62
QY 62 TDDGATDLDLGHYERFLKTYTKNNFTTGQYVEQVLNENKGDYLGATVQVYPIHTDS 121
DB 63 TDDGATDLDLGHYERFLKTYTKNNFTTGQYVEQVLNENKGDYLGATVQVYPIHTDS 122
QY 122 IKRRVYSEAGQVALIEVGTVGDIESTLPLETTIRKQGVGRDALFIHTLVPIYKS 181

```

```

DB 123 IKRRVYSEAGQVALIEVGTVGDIESTLPLETTIRKQGVGRDALFIHTLVPIYKS 182
QY 182 AGELTKPTQSHSVKELRTIGIOPDILICRSEOPIPASERKIALFTYVAKVISAIDND 241
DB 183 AGEVTKPTQSHSVKELRTIGIOPDILICRSDRIPANERKIALFTYVAKVISAIDND 242
QY 242 TYIRPILRLRBOGLDLDVVDQLRLDVPADLSAMEKYVDGLTHPTDEVSIALVGYVDHT 301
DB 243 STYKIPALRSQGLDELIVKFRSLTCREADLSSEWENVIYOGANRNGEVIAMVKIYELP 302
QY 302 DAVKSYNALIHAGIHRHKVQISYIDSETEAEAGTAKVNDALVPGGGEVGECKI 361
DB 303 DAVKSYNALIHAGIHRHKVQISYIDSETEAEAGTAKVNDALVPGGGEVGECKI 362
QY 362 STYVARENKIPYIGICIGMOSAVIEFARNVGLGASHTEFLPKSPHPVIGLITEMWDE 421
DB 363 LAAYVARENKIPYIGICIGMOSAVIEFARNVGLGASHTEFLPKSPHPVIGLITEMWDE 422
QY 422 AGEVLTRDSDLDGCTMELGAKQCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKOLE 481
DB 423 EGVNEORREHSDLDGCTMELGAKQCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKOLE 482
QY 482 AAGMKSQKSLDGRVLEIIELEPHPMFLACQFHPBETSTPRNGHALFSGFEAAAKH 539
DB 483 QAGLVPSGLSDRKLVEMLIELEPHPMFLACQFHPBETSTPRNGHALFSGFEAAAKH 540

```

RESULT 3

```

Q8GE64 PRELIMINARY; PRT; 498 AA.
ID Q8GE64
AC Q8GE64;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE CTP synthetase (fragment).
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;
OC Aeromonadaceae; Aeromonas.
OX NCBI_TaxID=644;
RN (1)
RP SEQUENCE FROM N.A.
RA Sha J., Galindo C.L., Pancholi V., Popov V.L., Chopra A.K.;
RT "Identification of a potentially New Virulence Gene Enolase that is
RT differentially expressed under in vitro versus in vivo growth
RT conditions in Aeromonas hydrophila.";
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY141757; AAN28925.1; -.
FT NON TER
SQ SEQUENCE 498 AA; 55079 MW; D9C6FC9228FBC80 CRC64;

```

Query Match 62.3%; Score 1738; DB 2; Length 498;
 Best Local Similarity 66.1%; Pred. No. 1.8e-112;
 Matches 326; Conservative 70; Mismatches 97; Indels 0; Gaps 0;

```

QY 47 DPGTMSFGQGEVYVTEEDGATDLDLGHYERFLKTYTKNNFTTGQYVEQVLNENKGD 106
DB 1 DPGTMSFGQGEVYVTEEDGATDLDLGHYERFLKTYTKNNFTTGQYVEQVLNENKGD 60
QY 107 YLGATVQYPIHTDSIKRRVYSEAGQVALIEVGTVGDIESTLPLETTIRKQGVGRDAL 166
DB 107 YLGATVQYPIHTDSIKRRVYSEAGQVALIEVGTVGDIESTLPLETTIRKQGVGRDAL 166
QY 167 RALFIHTLVPIYKSAGELTKPTQSHSVKELRTIGIOPDILICRSEOPIPASERKIALF 226
DB 121 NAMFMHLLTVLEYLAAGVVKTKPTQSHSVKELRTIGIOPDILICRSDRIPANERKIALF 180
QY 227 TNVAEKAVTSAIDQTYIRPILRLRBOGLDLDVVDQLRLDVPADLSAMEKYVDGLTHPT 286
DB 181 CNVERAVTSKMDVDSYIKPALKQSLDSYITERRGIKEKADLSEWENVIYOGANRNGEVI 240
QY 287 DEVSAIYGVYVDQVYKSYNALIHAGIHRHKVQISYIDSETEAEAGTAKVNDALVPGGGEV 346
DB 241 AEVTIIGMVKVSLPDAYKSYNALIHAGIHRHKVQISYIDSETEAEAGTAKVNDALVPGGGEV 300

```

QY 347 LNRGSGEGRVEKISTVRARANKIPLYGICLQMSAVIEFARNVGLRGAHSTFLEK 406
 DB 301 LNRGSGEGRVEKISTVRARANKIPLYGICLQMSAVIEFARNVGLRGAHSTFLEK 360
 QY 407 SPHPVIGLITWMDDEAGELVTRDESDLGITWMLAQKCEKLSADLAFOLYQCDVTEH 466
 DB 361 CAYPVGLITWMDDEAGELVTRDESDLGITWMLAQKCEKLSADLAFOLYQCDVTEH 420
 QY 467 RHRVEYNNKLLPQIEAAGLKVITGLSADKLVLEIIPDHFWFAAQFHEFISTPBDGHA 526
 DB 421 RHRVEYNNKLLPQIEAAGLKVITGLSADKLVLEIIPDHFWFAAQFHEFISTPBDGHA 480
 QY 527 LPSGFEYAAAKH 539
 DB 481 LPSGFEYAAAKH 493

RESULT 4
 ID Q8D2K0 PRELIMINARY; PRT; 552 AA.
 AC Q8D2K0;
 DT 01-MAR-2003 (TEMBLrel. 23, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)
 DE 01-MAR-2003 (TEMBLrel. 23, last annotation update)
 GN Pyrg protein.
 OS Wigglesworthia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.
 NCBI_TaxID=164609;
 RN [1]
 RP MEDLINE=22297718; PubMed=12219091;
 RX MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashita A., Matanabe H., Oshima K., Shiba T., Hattori M.,
 "Genome sequence of the endocellular obligate symbiont of tsetse
 flies, Wigglesworthia glossinidia."
 RL Nat. Genet. 32:402-407(2002).
 DR EMBL; AB063522; BAC24500.1; -
 KW Complete proteome.
 SQ SEQUENCE 552 AA; 62336 MW; D7DC35A8FF91C568 CRC64;

Query Match 59.6%; Score 1662; DB 16; Length 552;
 Best Local Similarity 58.7%; Pred. No. 4,1e-107;
 Matches 320; Conservative 90; Mismatches 123; Indels 12; Gaps 2;

QY 4 EFITGGVSSIGKGIASLAIEDRGKLTITKLDPIYVNDPGTMSPFQHGCVF 63
 DB 11 YFIFGGVSSIGKGIASLAIEDRGKLTITKLDPIYVNDPGTMSPFQHGCVF 70
 QY 64 DGAETDLDLGHYERPLKTTMTKKNFTTGOVEVLNRKGDYLGATVOVIPHITDIK 123
 DB 71 DGAETDLDLGHYERPLKTTMTKKNFTTGOVEVLNRKGDYLGATVOVIPHITDIK 130
 QY 124 RRVYSABKQVALLIEVGSTVGDIESLPLETRQMGVLEGRDRLFLHTLTPYIKAG 183
 DB 131 KQINAGKFDLLEHIGSTVGDIESLPLETRQMGVLEGRDRLFLHTLTPYIKAG 190
 QY 184 ELKTPQHSVLELRTIGIOPDILICRSBOPIPASERRKIALFTVAEKAIVSADDTI 243
 DB 191 EYKTPQHSVLELRTIGIOPDILICRSBOPIPASERRKIALFTVAEKAIVSADDTI 250
 QY 244 YHPIPLLEGGIDDLVVDQLRLDVPADLSAEKVVDDLTPTDEVSIAVGVYDHTDA 303
 DB 251 YHPIPLLEGGIDDLVVDQLRLDVPADLSAEKVVDDLTPTDEVSIAVGVYDHTDA 310
 QY 304 YKSLNEALIHAGIHRHKVQIYISIDSETI-BAEGTAKLKNVDAILVPGSGEGRVGEKIST 363
 DB 311 YKSLNEALIHAGIHRHKVQIYISIDSETI-BAEGTAKLKNVDAILVPGSGEGRVGEKIST 370
 QY 364 VFAPEENKIPLYGICLQMSAVIEFARNVGLRGAHSTFLEKSPHPVIGLITWMDDEAG 423

DB 371 AKYSREKIPYFGICLQMSAVIEFARNVGLRGAHSTFLEKSPHPVIGLITWMDDEAG 430
 QY 424 ELVTRDESDLGITWMLAQKCEKLSADLAFOLYQCDVTEHRRRVEYNNKLLPQIEA 483
 DB 431 YLNKSNISNNVSTVRLAQKCEKLSADLAFOLYQCDVTEHRRRVEYNNKLLPQIEA 487
 QY 484 GKPSGKSLD-----RLVETIIPHPWFLAQCFHEFISTPBDGHA 536
 DB 488 --KPSGKSLD-----RLVETIIPHPWFLAQCFHEFISTPBDGHA 545
 QY 537 KHKQTQ 541
 DB 546 DFKK 550

RESULT 5
 ID Q8EM53 PRELIMINARY; PRT; 535 AA.
 AC Q8EM53;
 DT 01-MAR-2003 (TEMBLrel. 23, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)
 DE 01-MAR-2003 (TEMBLrel. 23, last annotation update)
 GN CTP synthase (UTP-ammonia ligase) (EC 6.3.4.2).
 OS Oceanobacillus iheyensis.
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
 NCBI_TaxID=162710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=HT831 / DSM 14371 / JCM 11309;
 MEDLINE=2220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.,
 "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments."
 RL Nucleic Acids Res. 30:3927-3935 (2002).
 DR EMBL; AP004603; BAC14963.1; -
 KW Ligase; Complete proteome.
 SQ SEQUENCE 535 AA; 59550 MW; 2EA980BADC745E0 CRC64;

Query Match 57.9%; Score 1614; DB 16; Length 535;
 Best Local Similarity 57.1%; Pred. No. 8,4e-104;
 Matches 310; Conservative 86; Mismatches 133; Indels 14; Gaps 4;

QY 1 MKTIFIFGGVSSIGKGIASLAIEDRGKLTITKLDPIYVNDPGTMSPFQHGCVF 60
 DB 1 MKTIFIFGGVSSIGKGIASLAIEDRGKLTITKLDPIYVNDPGTMSPFQHGCVF 60
 QY 61 VTODGAETDLDLGHYERPLKTTMTKKNFTTGOVEVLNRKGDYLGATVOVIPHITD 120
 DB 61 VTODGAETDLDLGHYERPLKTTMTKKNFTTGOVEVLNRKGDYLGATVOVIPHITD 120
 QY 121 EIKRVYSABKQVALLIEVGSTVGDIESLPLETRQMGVLEGRDRLFLHTLTPYIKAG 178
 DB 121 EIKRVYSABKQVALLIEVGSTVGDIESLPLETRQMGVLEGRDRLFLHTLTPYIKAG 180
 QY 179 IKSAGELKTPQHSVLELRTIGIOPDILICRSBOPIPASERRKIALFTVAEKAIVSADDTI 238
 DB 181 IKSAGELKTPQHSVLELRTIGIOPDILICRSBOPIPASERRKIALFTVAEKAIVSADDTI 240
 QY 239 DADITVYHPIPLLEGGIDDLVVDQLRLDVPADLSAEKVVDDLTPTDEVSIAVGVYDHTDA 298
 DB 241 DADITVYHPIPLLEGGIDDLVVDQLRLDVPADLSAEKVVDDLTPTDEVSIAVGVYDHTDA 300
 QY 299 DHTDAVYSABKQVALLIEVGSTVGDIESLPLETRQMGVLEGRDRLFLHTLTPYIKAG 357
 DB 301 DHTDAVYSABKQVALLIEVGSTVGDIESLPLETRQMGVLEGRDRLFLHTLTPYIKAG 360
 QY 358 EKGISTVRAPEENKIPLYGICLQMSAVIEFARNVGLRGAHSTFLEKSPHPVIGLITWMDDEAG 417
 DB 361 EKGISTVRAPEENKIPLYGICLQMSAVIEFARNVGLRGAHSTFLEKSPHPVIGLITWMDDEAG 420
 QY 418 WMDDEAGELVTRDESDLGITWMLAQKCEKLSADLAFOLYQCDVTEHRRRVEYNNKLLPQIEA 476

Db 421 QKQ-----IEDLGSTLRNGAYPCQLVDGTAKAAYDQADVIERRRHRREFNNVY 470
 Qy 477 LKOLEAAGKAFESKSLDGRVLEIIELPEHPMFLACQHPDEFTSPRNGHALFSGFVEAAA 536
 Db 471 RDMWAKGAFVSGTSPDGRVLEIIKQHPMFLVACQHPDEFTSRPTAQSILFKFGIASV 530
 Qy 537 KHK 539
 Db 531 DTR 533

RESULT 6

ID	08G0G1	PRELIMINARY;	PRT;	542 AA.
AC	08G0G1			
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, last sequence update)		
DT	01-MAR-2003	(TrEMBLrel. 23, last annotation update)		
DE	CRP synthase.			
GN	PRNG OR BR1134.			
OS	Brucella suis.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Brucellaceae; Brucella.			
OX	NCBI_TaxID=29461;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1330 / Biovar 1;			
RX	MEDLINE=22247741; PubMed=12271122;			
RA	Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,			
RA	Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beaman M.J.,			
RA	Daugherty S.C., Debroy K.T., Durkin A.S., Kolonay J.F., Madupu R.,			
RA	Nelson W.C., Ayodele I.B., Krulul M., Shetty J., Malek U., Van Aken S.E.,			
RA	Riedmiller S., Tetreulin H., Gill S.R., White O., Salzberg S.L.,			
RA	Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;			
RT	"The Brucella suis genome reveals fundamental similarities between			
RT	animal and plant pathogens and symbionts."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).			
DR	EMBL; AEO141414; AAN30054.1;			
DR	TIGR; BR1134;			
KM	Complete proteome.			
SO	SEQUENCE	542 AA;	60058 MW;	B88F10406370B7B0 CRC64;

Query Match	56.9%;	Score 1585;	DB 16;	Length 542;
Best Local Similarity	58.1%;	Pred. No. 8.9e-102;		

Matches	312;	Conservative	78;	Mismatches	143;	Indels	4;	Gaps	4;
---------	------	--------------	-----	------------	------	--------	----	------	----

Qy	61	VTEDEGAFETDLDLGHYERFLKTTMTKNNFTTQGVLENERKGDYAGATVQYIPHIT	120
Dd	1	MARVVFITGVGVSSLGKGTIAAALALALCARGYRVRIRKIDLYLVNDPGTISPOHBEVF	60
Dd	61	VTDGAFETDLDLGHYERFLTRPANOQDNITTRIRIRNIEKERGCDYAGATVQYIPHITD	120
Qy	121	EIKRPRYAESAKGDVALIEVGTVGDISSLPRLEITTRQNGVGLGDRALFTHLTLVPIK	180
Dd	121	EIKRPFVLEGNEDDFVLCIEIGTVGDIEMAPLEIRQLGHELPRTGTAIVHTLTMPIR	180
Qy	181	SAGELTKTPROSHVKELRTGTGIDPLICRSQPRPASBRKRKILFTMVLEKATISIDA	240
Dd	181	ASGELTKTKPQSHVKELRSIGLAPDILVRAOREIPSSRRKLSIFCNVRESATQALDV	240
Qy	241	DTIYRIPLELRGSGLDLVVDOLRLD-VFAADLSAMEKYVGLTHTPDEVSIAYGVXD	299
Dd	241	ATITVPPIAVHKGDLSEVLNFAIGDIPAKPRMDMEVSHRLNPNBSBEVIAIVAGKYTG	300
Qy	300	HTDAYSLNEALIHAGITRHKVQSYDSIDSEITAEAGTAK-LKQNDALIVPGSGESGVE	358
Dd	301	LKDAYSLSLEALHGGILANKVKYKNTLMEIEAQFSESEDDAPYLEKHGILVPGSGEGEAG	360
Qy	359	GKISTVRFARENKIPYLGTICMGQSAVIEFANVYVLEGASTFELPSRHPVGLITTEM	418

Db 361 GKILAKFAERKVPFFGFCFGMCAACI EAARNVGI EHAASSBEGP - TREPVGLMTEW 412

Qy 419 MDASGLVLRDSDSLDGMTR-LGAKCKLAKDSLAFOLYQCDVITREHRRHRYEPNNQYLK 478

Db 420 L-KGNLLEGRRAAGDGLGCMTRGATVAVLKEDSKIAQVYSGTDLHRRHRRHYEVNI DYKD 478

Qy 479 QLEAAQMKSGSGSLQGRVRLITELPEHWPPLAACQPHPEFTSPENGALISGFI EAA 555

Db 479 KLLDAAGINAGKSPDQVLPETVYVYADHWFTGVQHPRLSKRPEPEPRLPASTFEAA 555

RESULT 7

ID	08CN12	PRELIMINARY;	PRT;	535 AA.
AC	08CN12;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	CTP synthase.			
GN	SEI125.			
OS	Staphylococcus epidermidis.			
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.			
OX	NCBI_TaxID=1282;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 12228;			
RA	Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,			
RA	Chen Z., Wen Y.;			
RL	Submitted (NOV-2002) to the EMBL/genbank/DBJ databases.			
DR	EMBL: A8016749; AAO05324.1; -			
KM	Complete proteome.			
SO	SEQUENCE 535 AA; 60211 MW; 629A03087D4AB238 CEC64;			

Query Match	56.5%	Score 1575	DB 16	Length 535
Best Local Similarity	55.6%	Pred. No. 4.3e-101		
Matches 302	95	Mismatches 132	Indels 14	Gaps 4

1. התאחדות העובדים - התאחדות העובדים הכללית, שהוקמה ב-1946, הייתה אחת מהתאחדויות העובדים הראשונות בישראל. היא נוסדה כתוצאה מיזוג התאחדויות העובדים העבריים והערבים, ונחשבת לאחת מהתאחדויות העובדים החזקות ביותר בישראל. התאחדות העובדים נמנית בין התאחדויות העובדים המייסדות של מדינת ישראל, ונחשבת לאחת מהתאחדויות העובדים החזקות ביותר בישראל.

Qy	1	MTKFIITIGVAVSSJLQKGIJAASSLAIJEDGKLKTEIKPDIYLVDPGNTSPQHEVF	60
Dh	1	MTKFIITIGVGVVSSJLQKGIJTAASLRLKQKLVKTIQKPDYLVDPGNTSPQHEVF	60
Qy	61	VTEDEAGETLIDGHTERPELTKTMTKGNFTTGQYBOYANERKGDYIGATVOYIPIITD	120
Dh	61	VTDGAGETLIDGHTERPELIDINIKNTXSVTFGKYSHVLKKERGDYIGATVOYIPIITN	120
Qy	121	EIRARVYSAS--GGOVAIEVGFTVGNJESLPLETIRQMGVELGRDRAFLHETLVPY	176
Dh	121	EIERLLIAGESTNADVVTTEIGGTTGDIESIPELIRKQIRSDUGERNVNYCTLLPY	180
Qy	179	IKSAGELKTKPTQHSYKELRTIGIOPDILICRSEOPIPASERRKIALFTVAAEKAVISAI	238
Dh	181	IKAGEMKTKPTQHSYKELRGIGIOPDILIVRTEYMTQDOKDIAIFCQIKESVTECR	240
Qy	239	DATYIRIPILLREOGLDIVDQIRLVLP--AALISAMEKVVLDTHPTDEVSIAIYGY	297
Dh	241	DASLVEITPQLSKQMDIVQIRQILNAKTEQDDEKHLINLVNNDGKITIGLVGY	300
Qy	298	VDHITDAYKSLNAL-IAGIHTRHKVQSYIDSETTEAGT-AKLKAVDAILVPGFGBCG	356
Dh	301	VSLQDVAIVSVESLKNHAGYPFKQDVVVMIDISEVNDENVEAYISVDGILVPGFGFRA	360
Qy	357	VSEKISTVPAEENKIPIYIGICLQMSAVIEPAAVNGVLBCASHTEPLPKSPHNVGLIT	416
Dh	361	SEKIDAIIVAAENNIPEFGICLQMLATVFAAHVAVGIBCAHSAELDPESTPIYIDLLP	420
Qy	417	EWMDAEGELVTRDESDIAGTWRGAKQCRKAKADSLAPOLYOKUVITERHHRYEENNOY	476
Dh	421	EQRD-----IEDAGTIRGLVGYCHIKKEGTLEKILNKNDIEERHRYEENNER	470
Qy	477	LKQLEAGKKFSGSKLDGRIVETIELPEHPWFLACQHPETSTSPRNGALVSGFYAAA	536
Dh	471	REOLBSNGVBSGTSDDRLVETIELPKNDFTIACQHPHEFLSRNRPQIPLASFBAAL	530

QY 537 RHK 539
Db 531 NYQ 533

RESULT 8
QBDXYO PRELIMINARY; PRT; 535 AA.

ID QBDXYO
AC QBDXYO;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE CTP synthetase (EC 6.3.4.2).
GN PYRG OR SPR0438.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus;
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszczyk L.C., Bunge S.,
DeHoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
Gilmour R., Glass J.S., Khaja H., Kraft A.R., Lagace R.E.,
LeBlanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,
McAhen S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
Zook C.A., Balz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatard P.L.,
RA Glass J.I.,
RT Genome of the bacterium Streptococcus pneumoniae strain R6.
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AF008423; AAK9242.1; -
KM Ligase; Complete proteome.
SQ SEQUENCE 535 AA; 59256 MW; 7BC6C267AF74C12 CRC64;

Query Match 56.0%; Score 1561; DB 16; Length 535;
Best Local Similarity 54.3%; Pred. No. 4; le-100;
Matches 294; Conservative 108; Mismatches 125; Indels 14; Gaps 4;

2 TKEFTFGVSSIGKGIASSLAIEDRGKATTTKLDPIYINVDPTGMSPPQHGVEY 61
3 TKITFTGCVSSIGKGIASSLAIEDRGKATTTKLDPIYINVDPTGMSPPQHGVEY 62
62 TTDGAEETDLDGHERFLTKTNNFTGQYEVYLRNRRKDYLGATVQYIPIHTD 121
63 TTDGAEETDLDGHERFLTKTNNFTGQYEVYLRNRRKDYLGATVQYIPIHTD 122
122 IKRRVYSAEGK--EKGVALIEVGTVGDIESTLPLETTRKMGVLEGRDALFIHLEVPY 179
123 IKRRVYSAEGK--EKGVALIEVGTVGDIESTLPLETTRKMGVLEGRDALFIHLEVPY 182
180 KSAEGLTKTPQHSVSKELRTIGIOPDILCRSEOPIPASERRKIALFTNVAEKAVISAID 239
183 KAAGEMTKTPQHSVSKELRTIGIOPDILCRSEOPIPASERRKIALFTNVAEKAVISAID 242
240 ADITTYRPLILREGDLDLVVDQRLDVPADLSAEKGVYDGLTHPTDVSIAIVGKYD 293
243 VEHLYQIPNLQKQGMQIVCDHLKLDAPADMTESAMVDKWNKKQKSLVKEYE 302
300 HTDAYSLSNLEALHAGIHERHKKVQISYIDSETTEAGTAK-LKNVADALIVPGSGGEGYV 358
303 LQDAIYLSVVALHSGVYNDVEKIMVNNNDVTANVALLSDAGIIVPGSGGEGYV 362
359 GKISTVAPAEKNTPIYIGICLQMSAVTEPRANVVGLEGAHSTFLPKSDHPYIGLITE 418
363 GKIQARVARENDVPMGLQWQICIFARHVLGEGANSALAEKTYPIIDIMROQ 422
419 MDAEGLVTRDEDSDLGTRMLGAQKRLKADSLAFOLY-QKDVYTERRRHRYEFNNQYL 477
423 ID-----IDMGCTIRLGLIYPSKLRGSKLAALAHNCEVQRRHRHRYEFNNQY 472
478 KOLEAAGMKFSKSLDGLVLEIIEPHEPWFLLACQHPHEFTSTPRNGHALPSGVEAAK 537

Db 473 EQFEAAGFYFSSGVSPDNRLVLEIIEPHEPWFLLACQHPHEFTSTPRNGHALPSGVEAAK 532
QY 538 H 538
Db 533 N 533

RESULT 9
QBDXYO PRELIMINARY; PRT; 543 AA.

ID QBDXYO
AC QBDXYO;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE CTP synthetase.
GN PYRG OR TLD0768.
OS Synecchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BP-1; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Saeamoto S.,
Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
RT Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005371; BAC08319.1; -
KM Complete proteome.
SQ SEQUENCE 543 AA; 60484 MW; 0884DA82309EFE40 CRC64;

Query Match 55.8%; Score 1555.5; DB 16; Length 543;
Best Local Similarity 57.0%; Pred. No. 1e-99;
Matches 310; Conservative 80; Mismatches 141; Indels 13; Gaps 3;

1 MTKFTFGVSSIGKGIASSLAIEDRGKATTTKLDPIYINVDPTGMSPPQHGVEY 60
1 MKITFTGCVSSIGKGIASSLAIEDRGKATTTKLDPIYINVDPTGMSPPQHGVEY 60
61 VTDDGAEETDLDGHERFLTKTNNFTGQYEVYLRNRRKDYLGATVQYIPIHTD 120
62 VTDDGAEETDLDGHERFLTKTNNFTGQYEVYLRNRRKDYLGATVQYIPIHTD 120
121 EIKRRVYSAEGK--DVALIEVGTVGDIESTLPLETTRKMGVLEGRDALFIHLEVPY 178
122 EIKRRVYSAEGK--DVALIEVGTVGDIESTLPLETTRKMGVLEGRDALFIHLEVPY 180
179 KSAEGLTKTPQHSVSKELRTIGIOPDILCRSEOPIPASERRKIALFTNVAEKAVISAID 238
181 KSAEGLTKTPQHSVSKELRTIGIOPDILCRSEOPIPASERRKIALFTNVAEKAVISAID 240
239 DADITTYRPLILREGDLDLVVDQRLDVPADLSAEKGVYDGLTHPTDVSIAIVGKYD 298
241 DKSITVAPAEKNTPIYIGICLQMSAVTEPRANVVGLEGAHSTFLPKSDHPYIGLITE 300
299 HTDAYSLSNLEALHAGIHERHKKVQISYIDSETTEAGTAK-ANKVADALIVPGSGGEGYV 357
301 RLSDAIYLSVVALHSGVYNDVEKIMVNNNDVTANVALLSDAGIIVPGSGGEGYV 360
358 GKISTVAPAEKNTPIYIGICLQMSAVTEPRANVVGLEGAHSTFLPKSDHPYIGLITE 417
361 EKIQAIVYARENDVPMGLQWQICIFARHVLGEGANSALAEKTYPIIDIMROQ 420
418 MDAEGLVTRDEDSDLGTRMLGAQKRLKADSLAFOLY-QKDVYTERRRHRYEFNNQYL 477
421 QDDIV-----DAGTIRLGLIYPSKLRGSKLAALAHNCEVQRRHRHRYEFNNQY 470
478 KOLEAAGMKFSKSLDGLVLEIIEPHEPWFLLACQHPHEFTSTPRNGHALPSGVEAAK 537

DB 471 NLEFETGYQITGTS PDGRVLEIIEYPAHFPIAVQHFHRSRPNAPHPLFYGLLAAAK 530
 QY 538 HKTQ 541
 DB 531 NSNR 534

RESULT 10

Q8E7P8 PRELIMINARY; PRT: 534 AA.
 AC Q8E7P8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN G8S0106.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=216495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MEM316 / Serotype III;
 RX MEDLINE=22242508; PubMed=12354221;
 RA Glaeser P., Rustiok C., Buchrieser C., Chevalier F., Frangeul L.,
 RA Msadek T., Zouine M., Couve E., Tallon L., Poyart C., Titeu-Cuot P.,
 RA Kunst F.;
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease."
 RL Mol. Microbiol. 45:1499-1513(2002).
 DR EMBL: AL766843; CAD45751.1; -;
 DR Sagalistic; G8S0106; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 534 AA; 59294 MW; 0CE1CC88511EDB5 CRC64;

Query Match 55.0%; Score 1533; DB 16; Length 534;
 Best Local Similarity 53.1%; Pred. No. 3.6e-98;
 Matches 288; Conservative 108; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEVF 60
 DB 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEVF 60
 QY 61 VTEHGAETDLDLGHYERFLKTKMTKKNFTTGOVYBOYLREKGDYLGATVOYPIHTD 120
 DB 61 VTEHGAETDLDLGHYERFLKTKMTKKNFTTGOVYBOYLREKGDYLGATVOYPIHTD 120
 QY 121 EIKRYYESA--EGKVALIEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 178
 DB 121 ALKEKIKRAATTDSVITTEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 180
 QY 179 IKSAGELKTRTQSHSVKELRTIGIPDILICRSEQPIPAERKIALFTNVAEKAVISAI 238
 DB 179 IKSAGELKTRTQSHSVKELRTIGIPDILICRSEQPIPAERKIALFTNVAEKAVISAI 240
 QY 239 DADTYIRIFELLREOGDLDLVVDQLRDVPAADLSAEKVDQLTPTPTVEISAIYKRY 298
 DB 239 DADTYIRIFELLREOGDLDLVVDQLRDVPAADLSAEKVDQLTPTPTVEISAIYKRY 300
 QY 241 DVDHITQIPLNMQQKNDQIVCDHLKLETPPADMTKMSAVDKVMLEKVKIALVGRY 300
 DB 241 DVDHITQIPLNMQQKNDQIVCDHLKLETPPADMTKMSAVDKVMLEKVKIALVGRY 300
 QY 299 DHTDAVYSLNALIHAGIHTRHAKVQIYSITSETIEAGTAKL-KNVDAIIVPGGFERGV 357
 DB 299 DHTDAVYSLNALIHAGIHTRHAKVQIYSITSETIEAGTAKL-KNVDAIIVPGGFERGV 360
 QY 358 EKGISTVREARENKIPLYIGICAGOSAVIEPAENVVGLGASTETELPKSPHPVIGLITE 417
 DB 358 EKGISTVREARENKIPLYIGICAGOSAVIEPAENVVGLGASTETELPKSPHPVIGLITE 420
 QY 418 WMDDEGEIVTRDSDLLGGTMRIGAKCKLADSLAPQY-OKDYITTEHRRHRYEPNNQY 476
 DB 418 WMDDEGEIVTRDSDLLGGTMRIGAKCKLADSLAPQY-OKDYITTEHRRHRYEPNNQY 470
 QY 421 QID-----IEDMGTLRLGLYPCILKAGSRAAAAYNQAVQVQHRHRRYEPNTKF 470
 DB 421 QID-----IEDMGTLRLGLYPCILKAGSRAAAAYNQAVQVQHRHRRYEPNTKF 470
 QY 477 LKQLBAAGKFFSGKSLDGLVETIELPEHPMFLAOCFFHEFTSTPRNGHALSSGVLEAA 536
 DB 477 LKQLBAAGKFFSGKSLDGLVETIELPEHPMFLAOCFFHEFTSTPRNGHALSSGVLEAA 536

DB 471 RQGFPAAGVFSGVSPDNRMLMEVELPEKKPFVAQAQYHPELOSRPNHAEIYATVTAAY 530
 QY 537 KH 538
 DB 531 EN 532

RESULT 11

Q8E290 PRELIMINARY; PRT: 534 AA.
 AC Q8E290;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CTP synthase
 GN PYRG OR SAG0107.
 OS Streptococcus agalactiae (serotype VI).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=216466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=22222988; PubMed=12200547;
 RA Tettele H., Masiyann V., Cieslewicz M.J., Eisen J.A., Peterson S.,
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
 RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
 RA Radu D., Fedorova N.B., Scanlan D., Khoult H., Mulligan S.,
 RA Caray H.A., Cline R.T., Van Aken S.E., Gill J., Scarsell M., Mora M.,
 RA Jacobini E.T., Bretton C., Gall G., Mariani M., Vegni F., Malone D.,
 RA Rinaldo D., Rappelli R., Telford J.L., Kasper D.L., Grandi G.,
 RA Frazer C.M.;
 RT "Complete genome sequence and comparative genomic analysis of an
 RT emerging human pathogen, serotype V Streptococcus agalactiae."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
 DR EMBL: AE014196; AAM99015.1; -;
 DR TIGR: SAG0107; -;
 KW Complete proteome.
 SQ SEQUENCE 534 AA; 59296 MW; 542AD55C132C870 CRC64;

Query Match 54.9%; Score 1532; DB 16; Length 534;
 Best Local Similarity 53.0%; Pred. No. 4.2e-98;
 Matches 287; Conservative 109; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEVF 60
 DB 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEVF 60
 QY 61 VTEHGAETDLDLGHYERFLKTKMTKKNFTTGOVYBOYLREKGDYLGATVOYPIHTD 120
 DB 61 VTEHGAETDLDLGHYERFLKTKMTKKNFTTGOVYBOYLREKGDYLGATVOYPIHTD 120
 QY 121 EIKRYYESA--EGKVALIEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 178
 DB 121 ALKEKIKRAATTDSVITTEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 180
 QY 179 IKSAGELKTRTQSHSVKELRTIGIPDILICRSEQPIPAERKIALFTNVAEKAVISAI 238
 DB 179 IKSAGELKTRTQSHSVKELRTIGIPDILICRSEQPIPAERKIALFTNVAEKAVISAI 240
 QY 239 DADTYIRIFELLREOGDLDLVVDQLRDVPAADLSAEKVDQLTPTPTVEISAIYKRY 298
 DB 239 DADTYIRIFELLREOGDLDLVVDQLRDVPAADLSAEKVDQLTPTPTVEISAIYKRY 300
 QY 241 DVDHITQIPLNMQQKNDQIVCDHLKLETPPADMTKMSAVDKVMLEKVKIALVGRY 300
 DB 241 DVDHITQIPLNMQQKNDQIVCDHLKLETPPADMTKMSAVDKVMLEKVKIALVGRY 300
 QY 299 DHTDAVYSLNALIHAGIHTRHAKVQIYSITSETIEAGTAKL-KNVDAIIVPGGFERGV 357
 DB 299 DHTDAVYSLNALIHAGIHTRHAKVQIYSITSETIEAGTAKL-KNVDAIIVPGGFERGV 360
 QY 358 EKGISTVREARENKIPLYIGICAGOSAVIEPAENVVGLGASTETELPKSPHPVIGLITE 417
 DB 358 EKGISTVREARENKIPLYIGICAGOSAVIEPAENVVGLGASTETELPKSPHPVIGLITE 420
 QY 421 QID-----IEDMGTLRLGLYPCILKAGSRAAAAYNQAVQVQHRHRRYEPNTKF 470
 DB 421 QID-----IEDMGTLRLGLYPCILKAGSRAAAAYNQAVQVQHRHRRYEPNTKF 470
 QY 477 LKQLBAAGKFFSGKSLDGLVETIELPEHPMFLAOCFFHEFTSTPRNGHALSSGVLEAA 536
 DB 477 LKQLBAAGKFFSGKSLDGLVETIELPEHPMFLAOCFFHEFTSTPRNGHALSSGVLEAA 536

QY 418 WMDAGELVTRDSDSLGTMRLGAOKCRLKADSLAFOLY-QKDYTERHRHRENNQY 476
 Db 421 QID-----IDMGGLRLGLYPCCKLSSGSRPAANNNVQRHRRHRENNQY 470
 QY 477 LKQLEAGMKFSKSLDGLRLVEILPEHFWFLACQHPHFTSTPRNGHALFSGFVEAAA 536
 Db 471 REQPEAGFVFSVSPDRLMEVVELPEKCFVAAQYHPELGRPHABELVTAFAVTA 530
 QY 537 KH 538
 Db 531 EN 532

RESULT 12

Q8F3J3 PRELIMINARY; PRT; 542 AA.
 ID Q8F3J3
 AC Q8F3J3
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, last annotation update)
 DE CTP synthase (EC 6.3.4.2).
 PYRG OR LA2409
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a;
 RL Ren S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE011408; AAN49608.1; -
 KW Ligase; Complete proteome.
 SQ SEQUENCE 542 AA; 61024 MW; 80BF9DBF864AF8A CRC64;

Query Match 54.9%; Score 1532; DB 16; Length 542;
 Best Local Similarity 53.4%; Pred. No. 4.3e-98;
 Matches 221; Conservative 99; Mismatches 141; Indels 14; Gaps 4;

QY 2 TKRIFITGVVSSLGKIAASLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 61
 Db 8 TKRIFITGVVSSLGKIAASLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 67
 QY 62 TEDGATDLDLGHYERFLKTTMTKKNFTTGQYEVOLRNERRKDYLGATVYIPIHIDE 121
 Db 68 TADGATDLDLGHYERFLKTTMTKKNFTTGQYEVOLRNERRKDYLGATVYIPIHIDE 127
 QY 122 IKRRVESA--EGKDVALLIEVGTVGDIISLPLETRIRQMGVELGRDALFIHLTLVPI 179
 Db 128 IKRRKATVARENPDIIVEIGTVGDIISLPLETRIRQMGVELGRDALFIHLTLVPI 187
 QY 180 KSAGELTKPTQHSVKELRTIGIOPDILICRSEQPIPAERKRIALFTNVAEKAVISAI 239
 Db 188 TAAEGAKTKPTQHSVKELRTIGIOPDILICRSEQPIPAERKRIALFTNVAEKAVISAI 247
 QY 240 ADT-ITRILLRBOGLDLDVVDQLRLDPADLSAMERVVDGLTHPTDVSIAIVGKYV 298
 Db 248 ISTSIEIKTKKEEDLVVLTNKMELRBSNFSFMDKMGVGLLTQKQVLAIVGKVI 307
 QY 299 DHTDAVKSINELIHAGIHTRRKVOISYIDSETIIEAGTAK-LKNVDALIVGGGGERGV 357
 Db 308 SIQDAVRSIYISLSHGIAHDTKEFIKYDPENLNDYSVELLKVHGILVGGGGERGI 367
 QY 358 EKKISTVAPAEKNTPIYLGICLQMSAVTEPARNVVGLGASHTEFLPKSPHPVIGLITE 417
 Db 368 EKKITAIQVARTNGPFIQICLQMSAVTEPARNVVGLGASHTEFLPKSPHPVIGLITE 427
 QY 418 WMDAGELVTRDSDSLGTMRLGAOKCRLKADSLAFOLY-QKDYTERHRHRENNQY 477
 Db 428 -----IDMGGLRLGLYPCCKLSSGSRPAANNNVQRHRRHRENNQY 477
 QY 478 LKQLEAGMKFSKSLDGLRLVEILPEHFWFLACQHPHFTSTPRNGHALFSGFVEAAA 537
 Db 478 KQYEEAGFVFSVSPDRLMEVVELPEKCFVAAQYHPELGRPHABELVTAFAVTA 537

QY 538 HKTQG 542
 Db 538 YSKKG 542

RESULT 13

Q8DMG1 PRELIMINARY; PRT; 536 AA.
 ID Q8DMG1
 AC Q8DMG1
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, last annotation update)
 DE CTP synthase (UTP-ammonia lyase) (EC 6.3.4.2).
 PYRG OR SMU.97.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1309;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN=0A159 / ATCC 700610 / Serotype C;
 RC MEDLINE=22295063; PubMed=12397186;
 RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 pathogen."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
 RL EMBL; AE014861; AAN57880.1; -
 KW Ligase; Complete proteome.
 SQ SEQUENCE 536 AA; 59491 MW; 1512BADCEBDF0446 CRC64;

Query Match 54.3%; Score 1514; DB 16; Length 536;
 Best Local Similarity 53.1%; Pred. No. 7.6e-97;
 Matches 288; Conservative 102; Mismatches 138; Indels 14; Gaps 4;

QY 1 MTKRIFITGVVSSLGKIAASLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 60
 Db 1 MTKRIFITGVVSSLGKIAASLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 60
 QY 61 VTDDGATDLDLGHYERFLKTTMTKKNFTTGQYEVOLRNERRKDYLGATVYIPIHIDE 120
 Db 61 VTDDGATDLDLGHYERFLKTTMTKKNFTTGQYEVOLRNERRKDYLGATVYIPIHIDE 120
 QY 121 IKRRVESA--EGKDVALLIEVGTVGDIISLPLETRIRQMGVELGRDALFIHLTLVPI 178
 Db 121 IKRRKATVARENPDIIVEIGTVGDIISLPLETRIRQMGVELGRDALFIHLTLVPI 180
 QY 179 KSAGELTKPTQHSVKELRTIGIOPDILICRSEQPIPAERKRIALFTNVAEKAVISAI 238
 Db 181 TAAEGAKTKPTQHSVKELRTIGIOPDILICRSEQPIPAERKRIALFTNVAEKAVISAI 240
 QY 239 DADTIYRILLRBOGLDLDVVDQLRLDPADLSAMERVVDGLTHPTDVSIAIVGKYV 298
 Db 241 DVDDYQIPPLVQKQNDQIVCDHLKQDVPADMTSMAVDKNNLKKKTKKILVGRVY 300
 QY 299 DHTDAVKSINELIHAGIHTRRKVOISYIDSETIIEAGTAK-TAKLNVDALIVGGGGERGV 357
 Db 301 ELPDAVLSVLEALHSGVNDPTALIDMINKANRETPREVTDLDGADGIIIVDGGGHRGT 360
 QY 358 EKKISTVAPAEKNTPIYLGICLQMSAVTEPARNVVGLGASHTEFLPKSPHPVIGLITE 417
 Db 361 EKKITAIQVARTNGPFIQICLQMSAVTEPARNVVGLGASHTEFLPKSPHPVIGLITE 420
 QY 418 WMDAGELVTRDSDSLGTMRLGAOKCRLKADSLAFOLY-QKDYTERHRHRENNQY 476
 Db 421 QID-----IDMGGLRLGLYPCCKLSSGSRPAANNNVQRHRRHRENNQY 470
 QY 477 LKQLEAGMKFSKSLDGLRLVEILPEHFWFLACQHPHFTSTPRNGHALFSGFVEAAA 536
 Db 471 KQYEEAGFVFSVSPDRLMEVVELPEKCFVAAQYHPELGRPHABELVTAFAVTA 530

QY 537 KH 538
DB 531 EN 532

RESULT 14

Q8G5X7 PRELIMINARY; PRT; 553 AA.
AC 08G5X7;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE CTP synthase.
GN PYRG OR BL0874.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NC 2705;
RX MEDLINE=22324977; PubMed=12381787;
RA Pessi G., Zwaan M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RA "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AB014709; AAN24687.1; -
KW Complete proteome.
SQ SEQUENCE 553 AA; 60983 MW; 8C8EA4ABECCF48C0 CRC64;

Query Match 51.4%; Score 1432.5; DB 16; Length 553;
Best Local Similarity 52.3%; Pred. No. 3.7e-91;
Matches 285; Conservative 97; Mismatches 150; Indels 13; Gaps 6;

QY 1 MKKFFITGVSYSKGLAASSLAILEDRLKTTITLQDYINVDGTSPPFGHYF 60
DB 13 VTKH:FYVGVGVSSLSKGLTASSLRLNSRGIHVLQOGLDYINVDGTMNPFQHGYY 72
QY 61 VEDGAEETLDLGHYERFLKTKKNNFTTQGYEQLVNERKGDYLGATVQVPHITD 120
DB 73 VEDGAEETLDLGHYERFLDVLISQKANTTQIQVYRKRAHAYLGQCVQVPHITD 132
QY 121 EIKRRV-YESAGKQVALIEVGTVGDIISLPPEFTIRQVVELGRDALFIHLTVPI 179
DB 133 EIKSMRAQASDVVILIEIGTVGDIESQPLEAAREVRDLGPRCMFHVSLVPEYI 192
QY 180 KSAEGLKTPTOHVSXELRTIGIOPDILICRSEOPIPASERKIALFTNVAKAVISAID 239
DB 193 SAAHEIKTPTOHVSXVMTLQGLISPDALVNSDRPLNOSIKKISLMCDVADAGVNCVD 252
QY 240 ADTIYRIPLLIREQGLDVLVDQLRDVPAADLSAMEKVDGLTPTDEVSIAIVKTYD 299
DB 253 ABSIDVPEKILPEEGDVAVYRELGLPFHDVDMENADLLEVRHHPKHEVNIAIVGKIID 312
QY 300 HTDAYSLSNEALIHAGIHRHVOISYIDSEITE-ABG-TALKNVVDALVPGGGEQV 357
DB 313 LPDAVLSVEALKAGPANMAVNVKVAADACETTEGAALADLVVDGIVIPGGGIGAI 372
QY 358 ESKISTVPEARENKIPYIGICLQMSAVIEFARNVVGLEGASTFELKSPHPVIGLITE 417
DB 373 DCKIGLTKRARETKIPALGICIGISQSVIIEYRHVLGIEDANSSFEEDCANPIVATMBE 432
QY 418 WMD-EGAGLVTEDESDLGCTWELGQKCLAKADSLAQOLYQKDVITERRHRYEFNNQY 476
DB 433 QMDIYAGK-----GDMGHTWRIGSYPALEEGSLVLELGTTHVTERHRRREVNAY 485
QY 477 LKQLRAAGKPSGKSLDGLVLEIIELEPR--HPWFLACQFHPFTSTPRNGALPSGFYEA 534
DB 486 KDLREGGIRISGQSPDGLTFVELPDVYHPFYATQAHHPFKRPTKRPPLPAGLVKA 545
QY 535 AAHKH 539

DB 546 ALDHQ 550

RESULT 15

Q8GD26 PRELIMINARY; PRT; 468 AA.
AC 08GD26;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE CTP synthase (EC 6.3.4.2) (Fragment).
OS Helicobacterium mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacterium.
OX NCBI_TaxID=28064;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22337798; PubMed=12446909;
RA Raymond J., Zhaxybayeva O., Cogarten J.P., Gerdes S.Y.,
RA Blankenship R.E.;
RT "Whole-genome analysis of photosynthetic prokaryotes."
RL Science 298:1616-1620(2002).
RN (2)
RP SEQUENCE FROM N.A.
RA Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
RA Gerdes S., Kyriades N., Overbeek R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY142825; AAN87429.1; -
KW Ligase.
KM NON TER
SQ SEQUENCE 468 AA; 52317 MW; 51E548CCB7D98AAS CRC64;

Query Match 50.8%; Score 1416.5; DB 2; Length 468;
Best Local Similarity 57.7%; Pred. No. 3.7e-90;
Matches 276; Conservative 73; Mismatches 116; Indels 13; Gaps 3;

QY 28 LEDRGLKVTITKLDPIYINVDGTSPPQGEVFTEDGAEETLDLGHYERFLKTKTKN 87
DB 1 MSSRGLKQICQKFPYINIDPGTSPYQHGSEVFTDGLAEETLDLGHYERFVDVNLKAS 60
QY 88 NPTGQVYEQVLRNRKGDYLGATVQVPHITDEIKRVSESA--EKNVALLIEVGTVG 145
DB 61 NVTKGIWYSYISKRRGYSYGVQVPHITDEIKRVSESA--EKNVALLIEVGTVG 120
QY 146 DIESLPELETIRQVVELGRDALFIHLTVPIKSNAGELKTKTOHVSXELRTIGIOPD 205
DB 121 DIESLPELETIRQVVELGRDALFIHLTVPIKSNAGELKTKTOHVSXELRTIGIOPD 180
QY 206 ILICRSEOPIPASERKIALFTNVAKAVISALDADITIRIPLILREQGLDVLVDQLRL 265
DB 181 IIVCTERKAPQELERKIALPCDIDPAVAVIOWDAETIYVPLILKEELDVIDITKILK 240
QY 266 DVPAADLSAMEKVDGLTPTDEVSIAIVKTYVDHTDAYSLNEALIHAGIHRHKOIS 325
DB 241 TCGSPDLKDMENALVHKIKNKQVTTIGLVKTYVALPANYSVASLSHAGIHRHNAVKIK 300
QY 326 YIDSEITE-ABGTRAKLVNDALVPGGFEKGVGKISTYRFAARENKIPIYIGICLQMSA 384
DB 301 WISYADLEGAPPEPFLSDVDGIIIVPGGFGRIEGKISALRYARENKIPIYIGICLQMSA 360
QY 385 VIEFARNVVGLEGASTFELPKSPHPVIGLITEMDAGELVTRDESDIGTWRLEAOK 444
DB 361 VVEYGRVVLGMDANSEFSETTHPIYIDILPPQKV-----EKGQKRLGSIWA 410
QY 445 CRLKADSLAQOLYQKDVITERRHRYEFNNQYKQLAAAGKPSGKSLDGLVLEIIE 502
DB 411 CRVAKGTWDAVQDEYIEYERHRRYEFNNNYEALIEKAGIISGTSPPDGLVLEVAL 468

Search completed: January 29, 2004, 15:54:12
Job time : 42.5492 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 ; Search time 29.0679 Seconds
(without alignments)
1736.455 Million cell updates/sec

Title: US-09-941-947a-18
Perfect score: 1623
Sequence: 1 MOIVANPRGFCAGVDRAIE.....GIEKVFSPKELKHMQA 318

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1623	100.0	318	23	AAE22307
2	1623	100.0	318	23	AAU80334
3	996	61.4	316	24	ABP60423
4	975	60.1	322	24	ABP77632
5	969.5	59.7	349	20	AAU23260
6	777	47.9	350	22	AAU5857
7	755	46.5	325	22	AAU90889
8	755	46.5	325	22	AAU78920
9	743.5	45.8	329	22	AAU81220

10	733	45.2	335	22	AAU81137
11	663	40.9	310	20	AAU75672
12	663	40.9	310	23	ABP62010
13	663	40.9	310	23	AAU09434
14	658	40.5	352	23	ABP66088
15	657	40.5	307	20	AAU37259
16	634.5	39.1	293	22	AAU67955
17	598.5	36.9	539	22	AAU51849
18	443.5	27.3	331	23	ABU47969
19	305	18.8	464	21	AAU15159
20	300	18.5	380	21	AAU26263
21	300	18.5	380	21	AAU38001
22	300	18.5	414	21	AAU26262
23	300	18.5	414	21	AAU38000
24	300	18.5	466	21	AAU26261
25	300	18.5	466	21	AAU37999
26	259.5	16.0	146	22	ABU18339
27	244.5	15.1	197	23	ABU51526
28	230	14.2	452	23	ABP93369
29	194.5	12.0	192	18	AAU20838
30	192.5	11.9	182	18	AAU20153
31	115	7.1	1466	22	ABU18175
32	115	7.1	2458	22	ABU24622
33	112	6.9	1033	23	ABP66122
34	110	6.8	500	23	ABU48818
35	109	6.7	578	22	AAU81233
36	108.5	6.7	544	12	AAU13337
37	104.5	6.4	539	22	AAU18448
38	104	6.4	488	22	AAU26267
39	104	6.4	494	23	ABP40179
40	104	6.4	443	20	AAU33930
41	103	6.3	433	22	AAU35014
42	103	6.3	553	23	ABP73501
43	103	6.3	585	18	AAU26665
44	103	6.3	585	20	AAU73895
45	103	6.3	896	22	AAU61465

ALIGNMENTS

RESULT 1		
ID	AAE22307	standard; Protein; 318 AA.
AC	AAE22307;	
XX		
DT	25-JUL-2002	(first entry)
XX		
DE	Methylomonas 16a sp. dimethylallyl diphosphate (lytB) enzyme.	
XX		
KW	Carotenoid; isopentenyl pyrophosphate; anthraxanthin; diet;	
KW	anti-oxidant; steroid; flavour; fragrance; electro-optic application;	
KW	agriculture; enzyme; dimethylallyl diphosphate; lytB.	
XX		
OS	Methylomonas 16a sp.	
XX		
PN	MO200218617-A2.	
XX		
PD	07-MAR-2002.	
XX		
PE	04-SEP-2001; 2001MO-US27420.	
XX		
FR	01-SEP-2000; 2000US-229858P.	
XX		
PR	01-SEP-2000; 2000US-229907P.	
XX		
PA	(DUPO) DU PONT DE NEMOURS & CO E I.	
XX		
PI	Broszowski PC, Cheng Q, Dicosimo DJ, Koffas M, Miller BS;	
XX	Odum WJ, Piatagglio SK, Rouviere PE;	
XX	WPI; 2002-351711/38.	
DR	N-PSDB; AAD35505.	

Mycobacterium tube
Chlamydia pneumoniae
C. pneumoniae BVH
Chlamydia pneumoniae
Bifidobacterium 10
Amino acid sequenc
Propionibacterium
P. faecium isop
Listeria monocytog
A. palaeolina LTB
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Helicobacter pylor
Novel human diagn
H. pylori inner me
Novel human diagn
Novel human diagn
Bifidobacterium 10
Listeria monocytog
Mycobacterium tube
HypB protein. Chl
S. epidermidis ope
Staphylococcus epi
Consensus mino aci
Enterococcus faeca
Candida albicans e
Yeast checkpoint c
Yeast MEK3 protein
Propionibacterium

XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by
 PT using microorganisms having a nucleic acid molecule encoding enzymes in
 PT the carotenoid biosynthetic pathway and which metabolize single carbon
 PT substrates

PS Claim 47; Page 123-124; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.
 CC The method comprises a transformed metabolizing host cell, comprising
 CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
 CC encoding an enzyme in the carotenoid biosynthetic pathway, under the
 CC control of regulatory sequences, and contacting the host cell with carbon
 CC substrate to produce a carotenoid compound. The method is useful for
 CC producing carotenoid compounds such as anthraxanthin and astaxanthin, by
 CC using microorganisms having a nucleic acid molecule encoding enzymes in
 CC the carotenoid biosynthetic pathway and which metabolize single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Methylomonas 16a sp. dimethylallyl diphosphate (lytB) enzyme used
 CC in the invention.

XX Sequence 318 AA;

SQ Query Match 100.0%; Score 1623; DB 23; Length 318;

Best Local Similarity 100.0%; Pred. No. 2,3e-141; Mismatches 0; Gaps 0;

Matches 318; Conservative 0; Indels 0; Gaps 0;

QY 1 MOIVLANPFGCAGVDRAIEIVDQALFAPGAPIVRHEVHNRTVVDLKKQKAVFIEEL 60

DB 1 MOIVLANPFGCAGVDRAIEIVDQALFAPGAPIVRHEVHNRTVVDLKKQKAVFIEEL 60

QY 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQKREVLIG 120

DB 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQKREVLIG 120

QY 121 HAGHEVEGTMGQYEKTEGGGIVLVETPEDVRLKVNPNDLAVYTQTLSTMTDKMV 180

DB 121 HAGHEVEGTMGQYEKTEGGGIVLVETPEDVRLKVNPNDLAVYTQTLSTMTDKMV 180

QY 181 DALPEQPSIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240

DB 181 DALPEQPSIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240

QY 241 GKPAVLIIDYQDLKQDMLEGEIEVGVTAGASAPVAVQVVDOLKAWGGETTSVRENSGI 300

DB 241 GKPAVLIIDYQDLKQDMLEGEIEVGVTAGASAPVAVQVVDOLKAWGGETTSVRENSGI 300

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

PD 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US26852.

XX 01-SEP-2000; 2000US-229907P.

XX (DUFO) DU PONT DE NEMOURS & CO E. I.

XX Chang Q, Koffas M, Norton KC, Odom JM, Picatagisio SK, Rouviere PE;

XX Schenle A, Tomb J;

XX WPI; 2002-383051/41.

XX N-PSDB; ABR50994.

XX Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,

XX isolated from Methylomonas 16a, useful for the production of isoprenoid

XX compounds -

XX Claim 4; Page 84; 84pp; English.

CC The present invention relates to a new nucleic acid molecule encoding
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.
 CC The invention is useful for obtaining a nucleic acid molecule
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the
 CC microbial production of isoprenoid compounds. The molecules of the
 CC organism are also useful for regulating isoprenoid biosynthesis in an
 CC organism and for producing recombinant organisms for producing various
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,
 CC for the production of keratoids and their derivatives, isoprenoid
 CC intermediates, and as pure products useful as pigments, flavours and
 CC fragrances. The present amino acid sequence represents the Methylomonas
 CC 16a open reading frame 10 (ORF10) lytB protein of the invention, as
 CC described above. This sequence functions in the formation of IPP
 CC (isopentenyl diphosphate) and dimethylallyl diphosphate.

XX Sequence 318 AA;

SQ Query Match 100.0%; Score 1623; DB 23; Length 318;

Best Local Similarity 100.0%; Pred. No. 2,3e-141; Mismatches 0; Gaps 0;

Matches 318; Conservative 0; Indels 0; Gaps 0;

QY 1 MOIVLANPFGCAGVDRAIEIVDQALFAPGAPIVRHEVHNRTVVDLKKQKAVFIEEL 60

DB 1 MOIVLANPFGCAGVDRAIEIVDQALFAPGAPIVRHEVHNRTVVDLKKQKAVFIEEL 60

QY 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQKREVLIG 120

DB 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQKREVLIG 120

QY 121 HAGHEVEGTMGQYEKTEGGGIVLVETPEDVRLKVNPNDLAVYTQTLSTMTDKMV 180

DB 121 HAGHEVEGTMGQYEKTEGGGIVLVETPEDVRLKVNPNDLAVYTQTLSTMTDKMV 180

QY 181 DALPEQPSIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240

DB 181 DALPEQPSIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240

QY 241 GKPAVLIIDYQDLKQDMLEGEIEVGVTAGASAPVAVQVVDOLKAWGGETTSVRENSGI 300

DB 241 GKPAVLIIDYQDLKQDMLEGEIEVGVTAGASAPVAVQVVDOLKAWGGETTSVRENSGI 300

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

QY 301 EEKVVFSPKELKKMOA 318

DB 301 EEKVVFSPKELKKMOA 318

XX Escherichia coli lytB protein SEQ ID NO 2.

XX Escherichia coli; mevalonate-independent isoprenoid synthesis pathway;
 XX MEP-way; Gcgs; LytB; T cell; asthma; Crohn's disease; multiple sclerosis;
 XX autoimmune disease; osteoporosis; tumor; antiasthmatic; antitumor;
 XX antiinflammatory; neuroprotective; immunosuppressive; antiasthmatic;
 XX osteopathic; enzyme.

OS Escherichia coli.

XX DE10119905-AL.

XX 24-OCT-2002.

XX 23-APR-2001; 2001DE-1019905.

XX 23-APR-2001; 2001DE-1019905.

XX (JOMA-) JOMAA PHARMARKA GMBH.

XX Jomaa H, Eberl M, Altincicek B;

XX MPI: 2003-113392/11.

XX N-PSDB; AB259204.

PT Enriching intermediates in the mevalonate-independent pathway of
 PT isoprenoid synthesis, useful for therapeutic activation of T cells,
 PT comprises altering enzymatic activity in the pathway

PS Disclosure; Page 6-8; 10pp; German.

CC The invention relates to enriching intermediates (A) in the
 CC mevalonate-independent isoprenoid synthesis pathway (MEP-way) comprising
 CC deleting, inactivating or otherwise altering a gene (i) in the pathway,
 CC in a cell or organism, so that the enzymatic activity of the product of
 CC (i) is reduced or made non-natural. The method is used for production of
 CC (A), especially substrates of the Gcgs and LytB enzymes that activate
 CC gamma/delta T cells. (A) and their derivatives are useful for:
 CC (i) determining activity of Gcgs and LytB, e.g. to identify their
 CC inhibitors;
 CC (ii) to activate gamma/delta T cells; and
 CC (iii) as pharmaceuticals. Dead or live cells or organisms enriched in (A)
 CC can be used similarly for treatment, in humans or animals, of asthma,
 CC Crohn's disease, ulcerative colitis, multiple sclerosis, chronic
 CC bronchitis, (auto)immune diseases, allergies, bone diseases and
 CC osteoporosis, also a wide variety of other diseases and for improving the
 CC immune response against tumors. The present sequence is that of the
 CC Escherichia coli LytB protein, inactivated in examples of the invention.

XX Sequence 316 AA;

Query Match 61.4%; Score 996; DB 24; Length 316;

Best Local Similarity 61.3%; Pred. No. 1.8e-83;

Matches 192; Conservative 59; Mismatches 58; Indels 4; Gaps 2;

1 MOTLVANPBGFCAGVDRALFIVDOALEAGAPLYVHEVYHNRVTYVDGLKQKGAAPFIEEL 60

1 MOILLANPBGFCAGVDRALFIVDOALEAGAPLYVHEVYHNRVTYVDGLKQKGAAPFIEEL 60

61 SDVPVSYLFFSAHGVSKVQOEAERQLTVPDAPCLVTXKVMQYAKAKAGREVILLG 120

61 SEVPPDAILIFSAHGVSKVQOEAERQLTVPDAPCLVTXKVMQYAKAKAGREVILLG 120

121 HGHPEVEGTMGQYKCTGGGIIYVETPEDYVNLKVNPNNDLAYVYQTLLSMTDTKVMV 180

121 HGHPEVEGTMGQYKCTGGGIIYVETPEDYVNLKVNPNNDLAYVYQTLLSMTDTKVMV 180

181 DALREGFPSIKQKODICATONRODAVHDLAKISDLILVWSPSSNSNRLEIAVOL 240

179 DALRKRFPIVGRKODICATONRODAVHDLAKISDLILVWSPSSNSNRLEIAVOL 240

241 GKPAYLIDITYDLKODMLGIEVGVYTAGASAPBVLVQEVIDQLAMGSETTSVRENSGI 300

DB 239 GKRAFLIDADIDQEBWKEVKCVTAGASAPOLIVGVYARLQQLGGGAIPLE--GR 296

QY 301 EKVYPSIPEKELK 313

DB 297 ENIVFEVPEKELR 309

RESULT 4

ABP77632

ID ABP77632 standard; Protein; 322 AA.

AC ABP77632;

DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 1794.

XX Antibacterial; infection; vaccine; gene therapy.

XX Neisseria gonorrhoeae.

XX WO200279243-A2.

XX 10-OCT-2002.

XX 12-FEB-2002; 2002WO-1B02069.

XX 12-FEB-2001; 2001GB-0003424.

XX (CHIR-) CHIRON SPA.

XX Fontana MR, Pizza M, Masiagnani V, Monaci B;

XX MPI: 2003-058415/05.

XX N-PSDB; AB238602.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a

XX medicament for treating or preventing N. gonorrhoeae infection

XX Disclosure; Page 317; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.

XX Also disclosed are the nucleic acid molecules encoding the proteins and

XX antibodies that specifically bind to the proteins. The composition

XX comprising the protein, nucleic acid or antibody is useful for the

XX manufacture of a medicament for treating or preventing N. gonorrhoeae

XX infection, this may be in the form of a vaccine or gene therapy.

XX Sequences given in records ABP76736-ABP81046 represent nucleic acid

XX molecules of the invention.

XX Sequence 322 AA;

Query Match 60.1%; Score 975; DB 24; Length 322;

Best Local Similarity 61.5%; Pred. No. 1.6e-81;

Matches 192; Conservative 48; Mismatches 66; Indels 6; Gaps 2;

3 IYLANPBGFCAGVDRALFIVDOALEAGAPLYVHEVYHNRVTYVDGLKQKGAAPFIEELSD 62

6 IYLANPBGFCAGVDRALFIVDOALEAGAPLYVHEVYHNRVTYVDGLKQKGAAPFIEELSD 65

63 VPVSYLFFSAHGVSKVQOEAERQLTVPDAPCLVTXKVMQYAKAKAGREVILLG 122

66 VPPDAILIFSAHGVSKVQOEAERQLTVPDAPCLVTXKVMQYAKAKAGREVILLG 125

123 HGHPEVEGTMGQYKCTGGGIIYVETPEDYVNLKVNPNNDLAYVYQTLLSMTDTKVMV 182

126 HGHPEVEGTMGQYKCTGGGIIYVETPEDYVNLKVNPNNDLAYVYQTLLSMTDTKVMV 182

183 DALREGFPSIKQKODICATONRODAVHDLAKISDLILVWSPSSNSNRLEIAVOL 242

182 LMARFPIVGRKODICATONRODAVHDLAKISDLILVWSPSSNSNRLEIAVOL 241

```

QY 243 PAYLIDVTQDLKQDMLEGIIEVGVTAGASAPFVLYOEVIIDOLKAMGGFTTSVRENSGIE 302
DB 242 DAYMDNABYIQRTEWEGSKSVGVITAGASAPFVLYOEVIIDOLKAMGGFTTSVRENSGIE 299
QY 303 KVFVSPISPELKK 314
DB 300 SIVFVLPKELRR 311

RESULT 5
AA129260
ID AAY29260 standard; Protein; 349 AA.
AC AAY29260;
XX
XX 25-OCT-1999 (first entry)
DT
XX Amino acid sequence of a virulence factor encoded by ORF4234c.
XX
XX Human pathogen; virulence polypeptide; virulence factor;
XX pathogenic infection; Pseudomonas aeruginosa infection.
XX
XX Pseudomonas aeruginosa.
XX
XX WO927129-A1.
XX
XX 03-JUN-1999.
XX
XX 25-NOV-1998; 98WC-US25247.
XX
XX 25-NOV-1997; 97US-0066517.
XX
XX (GENO) GEN HOSPITAL CORP.
XX
XX Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
XX Rahne IG, Tan M, Tsongalis J;
XX WPI; 1999-357851/30.
XX
XX Virulence factors useful in developing disease treatments
XX
XX Disclosure; Fig 4; 228pp; English.
XX
XX The present sequence represents a Pseudomonas aeruginosa polypeptide
XX sequence. P. aeruginosa is an opportunistic human pathogen present in
XX soil water and plants. The specification describes virulence polypeptides
XX and nucleic acid sequence encoding such polypeptides. These sequences
XX can be used to identify a compound which is capable of decreasing the
XX expression of a pathogenic virulence factor. Compounds that inhibit
XX the expression or activity of virulence factor polypeptides can be
XX used to treat pathogenic infections, especially where the infection
XX is a P. aeruginosa infection.
XX
XX note: the sequences given in the specification were poorly legible, and
XX in some instances assumptions were made as to the identity of the
XX residue; it is therefore possible that the sequence given below is
XX not entirely correct.
XX
XX Sequence 349 AA:
XX
XX Query Match 59.7%; Score 969.5; DB 20; Length 349;
XX Best Local Similarity 61.0%; Pred. No. 5.9e-81;
XX Matches 191; Conservative 47; Mismatches 72; Indels 3; Gaps 2;
XX
XX 1 MGIIVANRGCAGVDRALRIYDQALIEARGAPIVRHEVYNNRTVDGKQKGFIEEL 60
XX 36 MGIKAMPFGFCAGVDRALRIYDQALIEARGAPIVRHEVYNNRTVDGKQKGFIEEL 95
XX
XX 61 SDVPVSYLIFSAHGVSKVQDAEERQUTVPDACPVTYKVMQVANHAKQGREVILIG 120
XX 96 DQVPMNVITVIFSAHGVSKVQDAEERQUTVPDACPVTYKVMQVANHAKQGREVILIG 155
XX
XX 121 HAGHEVBTGMOQKCTBGGGIIYVETPEYDVRNKKVNPNDLAVYQTITSMDDTKMV 180
XX

```

```

DB 156 HGHPEVBTGMOQKCTBGGGIIYVETPEYDVRNKKVNPNDLAVYQTITSMDDTKMV 214
QY 181 DALRQPFSTIEQKKDDICVATQNRQDAFNDIACISLIIYVSGPNSNSRRLAIVQL 240
DB 215 DALRQPFSTIEQKKDDICVATQNRQDAFNDIACISLIIYVSGPNSNSRRLAIVQL 274
QY 241 GKPAYLIDVTQDLKQDMLEGIIEVGVTAGASAPFVLYOEVIIDOLKAMGGFTTSVRENSGIE 300
DB 275 GTPAYLIDGADMDQRCGFCGRIRIGITRAGASAPFVLYOEVIIDOLKAMGGFTTSVRENSGIE 332
QY 301 EKVVPISPELKK 313
DB 333 EKVVPISPELKK 345

RESULT 6
AAUS8857
ID AAUS8857 standard; Protein; 350 AA.
AC AAUS8857;
XX
XX 13-FEB-2002 (first entry)
DT
XX Propionibacterium acnes immunogenic protein #19753.
XX
XX DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX
XX 02-JUN-2000; 2000US-208841P.
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhargava A;
XX L'Amalbonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX
XX N-PSDB; AAS59596.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX Example 1; SEQ ID No 20052; 1069pp; English.
XX
XX Sequences AAU39105-AAU588017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX

```


PR 01-JUL-1999; 99DE-1030429.
 PR 01-JUL-1999; 99US-0142692.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031541.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032914.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99US-0151214.
 PR 31-AUG-1999; 99DE-1041382.

XX (BAD) BASF AG.

PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G, Lee H;
 PI Kim H;

XX WPI, 2001-061972/07.
 XX N-PSDB; AAF71033.

DR New isolated *Corynebacterium glutamicum* nucleic acid encoding a stress,
 PT tolerance or resistance protein, for production or modulation of
 PT production of fine chemicals, such as, e.g. amino acids, lipids,
 PT carbohydrates, or enzymes -

PS Claim 20; Page 249-250; 526pp; English.

XX AAF70984 to AAF71133 encode the *Corynebacterium glutamicum* stress,
 CC resistance, and tolerance (SRT) proteins given in AAB78871 to AAB79020.
 CC The C. glutamicum SRT genes (I) can be used in vectors (II) for
 CC expression in host cells and production of fine chemicals, such as, an
 CC organic acid, a proteinogenic or nonproteinogenic amino acid (preferred),
 CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a
 CC saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic
 CC compound, a vitamin, a cofactor, a polypeptide, or an enzyme. The fine
 CC chemical production can be modulated. The presence of (I) or the SRT
 CC proteins (III) encoded by them are used for diagnosing the presence
 CC or activity of *Corynebacterium diptheriae*. (II), (III), (III) and host
 CC cells containing them can be used to map the genomes of organisms related
 CC to C. glutamicum, to identify and localise C. glutamicum sequences of
 CC interest, in evolutionary studies, in determination of SRT protein
 CC regions required for function, in modulating the SRT protein activity,
 CC and in modulating the activity of an SRT pathway. (II) are used to permit
 CC C. glutamicum to survive in an environment that is normally
 CC environmentally or chemically hazardous to it. (I) and protein molecules
 CC encoded by it increase the survival of C. glutamicum to chemical and
 CC environmental hazards and provide a means for continued growth and
 CC multiplication in large scale fermentative growth conditions. By
 CC increasing the growth rate or maintaining a normal growth rate in poor or
 CC toxic conditions, the yield, production and/or efficiency or production
 CC of fine chemicals from a culture may be increased.

XX Sequence 325 AA;

Query Match 46.5%; Score 755; DB 22; Length 325;

Best Local Similarity 48.3%; Pred. No. 3.4e-61;

Matches 152; Conservative 67; Mismatches 86; Indels 10; Gaps 4;

QY 2 QIVANPRGFCAGVDAIETVDQIAIFGAPLYVHEVHNTVVDGKQGAFTIELS 61
 DB 15 KILLAAPRGYCAVRAVETVERALREYCAPTYVEKEIVHRYVDTTIAEKAIIFVNAAS 74
 QY 62 DVPVGSYLIFSAGVSKVEQOAEEROLTVFDATCELVTKVMQVAKIAKQREVTILGH 121
 DB 75 EAPEGANMVFSAHGSFVHEHAAKNITAKITDAACPLVTKHKEIQRPDKQGFILFIGH 134
 QY 122 AGHPEVSGTMQGYEKCTEGGAGIYVETPEDVNLK--VNNPDLAYVTQTLLSMTDTVM 179
 DB 135 EGHHEEVEGTMGHSEVET-----HIVDGVAGIATLPEFLNDEPNLIMLSQTTLSVDRWEI 189
 QY 180 VVALSEOPSIKEQKKDICAATONROPAVGHDLATISULIIVGSPNSNKRLEIAYQ 239
 DB 190 VVELKKEFPQLODPSPSDICATYONKQVAVKAIARCEIMLTIVGSSRNSSNRYLVEVAKQ 249

QY 240 LGKP-AYCIDFTYODLKQDWLRGIEVGTAGASAPVLYVQEVINDLKAWGETTSYRENS 298
 DB 250 NGADNAYIVDAKAKIDPAMFEGVETIGISSGASVPEILVQGVETRLAEFGTD--DVEEYV 307
 QY 299 GIERKVFYSIPKEIK 313
 DB 308 SAAEKIVPALPRVLR 322

RESULT 9

AA681220
 ID AAG81220 standard; Protein, 329 AA.

XX AAG81220;

AC AAG81220; (first entry)

DE Mycobacterium tuberculosis potential drug target protein SEQ ID 271.

XX Drug target; growth; organism viability; characterisation.

OS Mycobacterium tuberculosis.

XX WO200135317-A1.

XX 17-MAY-2001.

XX 13-NOV-2000; 2000WO-US31152.

XX 12-NOV-1999; 99US-0165086.

XX 12-NOV-1999; 99US-0165124.

XX 01-FEB-2000; 2000US-0179531.

XX (RBCG) UNIV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

DR WPI; 2001-329193/34.

DR N-PSDB; AAH52071.

PT Identifying nucleotide or polypeptide sequence for use as drug target.

PT involves providing algorithm that analyzes a functional relationship

PT between nucleotide or polypeptide sequences, and comparing the

PT sequences -

PS Disclosure; Page 187; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or

XX polypeptide sequence that may be a drug target, or essential for growth

XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

XX represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium

XX tuberculosis proteins which are potential drug targets. The DNA and

XX protein sequences are used to illustrate the method of the invention. The

XX method involves providing an unknown nucleotide or polypeptide sequences,

XX and comparing it to a number of sequences along with at least one

XX algorithm capable of analysing a functional relationship between

XX nucleotide and polypeptide sequences. The method is useful for

XX characterising the function of nucleic acids and polypeptides that may be

XX useful as a target for a drug or essential for the growth or viability of

XX an organism.

XX Sequence 329 AA;

Query Match 45.8%; Score 743.5; DB 22; Length 329;

Best Local Similarity 47.6%; Pred. No. 4e-60;

Matches 150; Conservative 60; Mismatches 94; Indels 11; Gaps 3;

QY 3 IVALNPRGFCAGVDAIETVDQIAIFGAPLYVHEVHNTVVDGKQGAFTIELSD 62
 DB 20 VILASPPSFCAGVRAIETVDAVAGVYVYKQIVHNTVVAALRDGAIFVBDLDE 79
 QY 63 V----PVGSYLIFSAGVSKVEQOAEEROLTVFDATCELVTKVMQVAKIAKQREVTIL 118

Db 80 IPDPPEGAVVPSAHVSPAVRAGADERGLGVADNCPVAVVHAARARQTVVF 139
 Qy 119 IGHAGHEVVGTMGVQVEKTEGGGIVLVEPPEYRUKXANPNDAVYTOTLISMTDFK 178
 Db 140 IGHAGHEVVGTMGVQVEKTEGGGIVLVEPPEYRUKXANPNDAVYTOTLISMTDFK 194
 Qy 179 MDVALREQPSIKQKDDICVATQNRQDAVHDLAKISDLILVVGSPNSNSNRLEIAV 238
 Db 195 VIDALARFPTLCQPSDEDICVATNRQRLQSGMGECDVIVIGSCNSNSRRLVELAQ 254
 Qy 239 QLGKPAVLLIDTYDLKQDMLEGVVGTAGASAPVTVQEVTDOLKANGGETTSVRENS 298
 Db 255 RSGTPAVLLIDGPDDISEMILSVSTIGVTAGASAPRLVGQVTDALRGYASIT--VYERS 312
 Qy 299 GIEKVFSPKELK 313
 Db 313 IATETVRFGLPKQVR 327

RESULT 10

AAG81137
 ID AAG81137 standard; Protein; 335 AA.
 AC AAG81137;

04-SEP-2001 (first entry)

Mycobacterium tuberculosis potential drug target protein SEQ ID 188.

Drug target; growth; organism viability; characterisation.

Mycobacterium tuberculosis.

MO200135317-A1.

17-MAY-2001.

13-NOV-2000; 2000MO-US31152.

12-NOV-1999; 99US-0165086.

12-NOV-1999; 99US-0165124.

01-FEB-2000; 2000US-0179531.

(RBCG) UNIV CALIFORNIA.

Eisenberg D, Rotstein SH, Marcotte EM;

WPI; 2001-329193/34.

N-PSDB; AAH51988.

Identifying nucleotide or polypeptide sequence for use as drug target,

involves providing algorithm that analyzes a functional relationship

between nucleotide or polypeptide sequences, and comparing the

sequences -

Disclosure; Page 165; 207pp; English.

This invention relates to a method for identifying a nucleotide or

polypeptide sequence that may be a drug target, or essential for growth

or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium

tuberculosis proteins which are potential drug targets. The DNA and

protein sequences are used to illustrate the method of the invention. The

method involves providing an unknown nucleotide or polypeptide sequences,

and comparing it to a number of sequences along with at least one

algorithm capable of analysing a functional relationship between

nucleotide and polypeptide sequences. The method is useful for

characterizing the function of nucleic acids and polypeptides that may be

useful as a target for a drug or essential for the growth or viability of

an organism.

Sequence 335 AA;

Query Match 45.2%; Score 733; DB 22; Length 335;
 Best Local Similarity 47.0%; Pred. No. 3,8e-59;
 Matches 147; Conservative 66; Mismatches 92; Indels 8; Gaps 3;
 Qy 2 QVLANPFGCAGVRAIEIVDQALEAGAPVYVHEVENVNTVVDLQKQGAFFLEIS 61
 Db 27 RLIALPFRVYCGVRAVETVRBALQKQGPVYVHEVENVNTVVDLQKQGAFFLEIS 86
 Qy 62 DVPVGSYLLIFSAGVSKVQQAERQLTVPDCTCPVTVKVMQVAKAKQGEVYLLIGH 121
 Db 87 QVPEGAIVFSAHGAAPVTHVSASERNQVTDATCPVTVKVMQVAKAKQGEVYLLIGH 146
 Qy 122 AHPPEVETMGQVEKTEGGGIVLVEPPEYRUKXANPNDAVYTOTLISMTDFK 181
 Db 147 EGHREVGTAGE-----APDHQVLDVGDVADQVTVREDEKVMVLSQTTLSVDEMEIVG 201
 Qy 182 ALRQPSIKQKDDICVATQNRQDAVHDLAKISDLILVVGSPNSNSNRLEIAVQK 241
 Db 202 RLRRRFPKQDPPSDICVATQNRQVAVKAMAPCELVIVVGSRNSNSVRLEVALGAG 261
 Qy 242 -KPAVLIDTYDLKQDMLEGVVGTAGASAPVTVQEVTDOLKANGGETTSVRENSGI 300
 Db 262 ARAALVDMADDDISAMLDGVTVGVTSASVPEVTVRGVLERLACGYDI--VQPVTTA 319
 Qy 301 BEKVFSPKELK 313
 Db 320 NETLVFALPRELR 332

RESULT 11

AAY35672
 ID AAY35672 standard; Protein; 310 AA.
 AC AAY35672;

13-SEP-1999 (first entry)

Chlamydia pneumoniae cellular envelope protein.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

vaccine; neutralising epitope.

Chlamydia pneumoniae.

WO9927105-A2.

03-JUN-1999.

20-NOV-1998; 98WO-IB01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(GBST) GENSET.

Genome sequence of Chlamydia pneumoniae

Page 1389; Disclosure; 1912pp; English.

AAY34584-Y35879 represent the proteins encoded by all the open reading

frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.

C. pneumoniae causes respiratory disease such as pneumonia and

bronchitis and is thought to be a contributing factor in heart

disease, sarcoidosis, sinusitis, purulent otitis media, erythema

nodosum or pharyngitis. The polypeptides encoded by the open reading

frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in

immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotides sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of *C. pneumoniae*.

XX Sequence 310 AA;

Query Match 40.9%; Score 663; DB 20; Length 310;

Best Local Similarity 45.0%; Pred. No. 9.9e-53;
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

CC The invention relates to novel isolated polypeptides from *Chlamydia*

CC pneumoniae. The polypeptides of the invention have antibiotic and

CC antibacterial activity. The polypeptide or composition of the invention

CC is useful for the prophylactic or therapeutic treatment of *Chlamydial*

CC bacterial infection (specifically those caused by *C. pneumoniae*,

CC *C. psittaci* or *C. trachomatis*), e.g. sinusitis, pharyngitis, bronchitis,

CC pneumonia, asthmatic bronchitis, adult-onset asthma, chronic

CC obstructive pulmonary disease (COPD), atherogenesis or atherosclerosis.

CC The sequences shown in ABP61993-ABP62011 represent proteins BVH-CPN1 to

CC BVH-CPN19 of *C. pneumoniae*.

XX Sequence 310 AA;

Query Match 40.9%; Score 663; DB 23; Length 310;

Best Local Similarity 45.0%; Pred. No. 9.9e-53;
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

CC The invention relates to novel isolated polypeptides from *Chlamydia*

CC pneumoniae. The polypeptides of the invention have antibiotic and

CC antibacterial activity. The polypeptide or composition of the invention

CC is useful for the prophylactic or therapeutic treatment of *Chlamydial*

CC bacterial infection (specifically those caused by *C. pneumoniae*,

CC *C. psittaci* or *C. trachomatis*), e.g. sinusitis, pharyngitis, bronchitis,

CC pneumonia, asthmatic bronchitis, adult-onset asthma, chronic

CC obstructive pulmonary disease (COPD), atherogenesis or atherosclerosis.

CC The sequences shown in ABP61993-ABP62011 represent proteins BVH-CPN1 to

CC BVH-CPN19 of *C. pneumoniae*.

XX Sequence 310 AA;

Query Match 40.9%; Score 663; DB 23; Length 310;

Best Local Similarity 45.0%; Pred. No. 9.9e-53;
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

CC The invention relates to novel isolated polypeptides from *Chlamydia*

CC pneumoniae. The polypeptides of the invention have antibiotic and

CC antibacterial activity. The polypeptide or composition of the invention

CC is useful for the prophylactic or therapeutic treatment of *Chlamydial*

CC bacterial infection (specifically those caused by *C. pneumoniae*,

CC *C. psittaci* or *C. trachomatis*), e.g. sinusitis, pharyngitis, bronchitis,

CC pneumonia, asthmatic bronchitis, adult-onset asthma, chronic

CC obstructive pulmonary disease (COPD), atherogenesis or atherosclerosis.

DE C. pneumoniae BVH-CPN18 from strain CWL-029.

KW CWL-029; BVH-CPN; antibiotic; antibacterial; Chlamydial infection;

KW sinusitis; pharyngitis; bronchitis; pneumonia; asthmatic bronchitis;

KW adult-onset asthma; chronic obstructive pulmonary disease; COPD;

KW atherogenesis; atherosclerosis.

OS Chlamydia pneumoniae.

PN EP1219635-A2.

PD 03-JUL-2002.

PF 21-DEC-2001; 2001EP-0130295.

PR 21-DEC-2000; 2000US-256941P.

PI (SHIR-) SHIRE BIOCHEM INC.

PI Couture F, Hamel J, Brodeur BR, Martin D;

PI WPI; 2002-530680/57.

PI N-PSDB; ABQ92504.

PS Claim 9; Figure 36; 122pp; English.

RESULT 13

AAU09434 standard; Protein; 310 AA.

AAU09434;

26-MAR-2002 (first entry)

Chlamydia pneumoniae metalloproteinase.

KW ATP binding cassette; secretory locus open reading frame; endopeptidase;

KW secretory locus ORF; protease; metalloproteinase; Clp protease Atrase;

KW Clp protease subunit; transglycolase/transpeptidase; ClpC protease;

KW thioredoxin; Chlamydia infection; antibacterial.

OS Chlamydia pneumoniae CWL029.

PN WO200185972-A2.

PD 15-NOV-2001.

PF 08-MAY-2001; 2001WO-CA00653.

PR 08-MAY-2000; 2000US-202672P.

PR 30-MAY-2000; 2000US-207852P.

PR 16-JUN-2000; 2000US-211796P.

PR 16-JUN-2000; 2000US-211797P.

PR 16-JUN-2000; 2000US-211798P.

The present invention relates to the isolation of Chlamydia pneumoniae strain CW029 genes and their encoded proteins. The genes of the invention encode an ATP binding cassette gene, a secretory locus open reading frame (ORF), an endopeptidase, a protease, a metalloprotease, Clp protease AtpAse, a Clp protease subunit, a transglycosylase/transpeptidase, a Ctpc protease, or thioredoxin. The genes of the invention can be used in a vector as a vaccine for the prevention and treatment of Chlamydia infections. Also described are B- and T-cell epitopes from the proteins of the invention which can be used as Chlamydia antigens. AAU09430-AAU09439 represent the proteins encoded by the C. pneumoniae genes (AA518750-AA518755) of the invention

Sequence 310 AA;
xx
xx

```

QY 2 QIVLANPGRFCAGVDRAIEIVDOALEAFGAP IYVREHVVHNRVTVDGLKQKGAVFTEELS 61
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 3 KILCPGPRGFCGCVVRALQVVEVALLEKMGAGI IYMHHEIVHRRVNRNALIRAKGAI FYEEELV 62
QY 62 DVPVSGSYLLFSAHGSKKEVOEAEEROLTVEDATCP IYVTKHMQVAAHAKQGEVLLIG 121
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 63 DVPBGRVIVYSAHGIPEPSVRAEAKRKRLID IALTCGLVYKHSAAHLYAKGKRIILLIG 122
QY 122 AGHPEVSGTGGYEKCTEGGGIYLVETPEDEVRLN KVNPNDLAVYOTLLSMYDTKVMYD 181
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 123 KKHVEYLTGVGVGVP-----HITVEKADVDEALPFSISDIPFLFYIQTLLSDVQAEIS 177
QY 182 ALREOPPSIKEKCKDICIYATONRODAVHDIAKISD ILIYVSGPNSNSNRLEIEIVQ 241
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 178 ALTKVPSIITLPSSICVYATTNQKALRSTLSVNRV VVYVGVDPNSNSNRLEIEVALRG 237
QY 242 KRAYLIDTYQDLKQOMLEGIEVGYTLAGASPELTVQ SVIQLKAMGERTSYRENSGLE 301
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 238 VPADLLINPEDIDTNIVNHSGLDIAMTAGASPELTVQ ACIRKL-----SSLIPELQVE 290
QY 302 -----EKVPSIPREKL 313
Db 291 NDIFAVEDVEVQLPRELR 308

```

KW rotavirus; food composition; pharmaceutical composition.
 XX
 OS Bifidobacterium longum.
 XX
 PN EP1227152-A1.
 XX
 PD 31-JUL-2002.
 XX
 PF 30-JAN-2001; 2001EP-0102050.
 XX
 PR 30-JAN-2001; 2001EP-0102050.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 XX
 DR WPI; 2002-668397/72.
 XX
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
 PT as a probe or primer for detecting and/or identifying Bifidobacterium
 PT longum in a biological sample -
 PT

Claim 3; SEQ ID 832; 80pp; English.

The present invention describes a polynucleotide (I) comprising a sequence of a *Bifidobacterium* genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABP65258 to ABP65354, ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrhetic and antibacterial activities, and can be used as an inhibitor of *Salmonella*. (II) (which is a probe) is useful for the detection and/or identification of *Bifidobacterium longum* in a biological sample. A carrier containing the lactic acid bacterium *Bifidobacterium longum* NCC3705 (NCIM 1-2618) can be used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formula, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the *Bifidobacterium* gene. ABQ81844 to ABQ81850 represent *Bifidobacterium* related nucleotide sequences given in the sequence listing from the present invention but not mentioned further within the specification.

N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office.

Sequence 352 AA;

Query Match	40.5%	Score 658	DB 23	Length 352
Best Local Similarity	43.8%	Pred. No. 3.5e-52		
Matches 145	Conservative	64	Mismatches 96	Indels 26
			Gaps 7	

```

      3 IYLANPGRFCAGVDRAI--EIVDQAIKAFG-----APIVVRHEVVHNRKTVDGLKQ 51
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

27 VLIADPRGFCAGVDRAILITVQTLIKAAEASGKRTREDGEPVYVRQIVHNKXVEDILAG 86

52 KGAVIBELSDVP-----VGSYLIFSNAGVSKETVQQAERQLTFPDATCPILVTKHNGV 100

107 AKHAKOGRFVIT,IGHAGHREVECTMGZOVEKCTTEGGCTVY,VRTPDREYNT,KYANNDNDI,AYU 160
8/ QGAVEVQELMADI,PDARHQAGIEPVVFSABGVSP,VVKAHEAKRGHNVDAIC,PVGNVHREYV 148

147 LRFRVEGEYELVIYIGKRGHDAVGWESP-----FVHLIEHESDVSLSLCAEDTKYVLL 2015

167 TOTITSMITDKVMDBALREOPPSIKEQKDDICVATQNRQDAVEDLAKISDILVVGSPN 226

202 SQTLSVDEFTALTIATLAKAKPFMIQEPPSSDCVATSNQGAAYKLVAGQSPCCVAVIGSAN 2611

227 S8NSNRLEIAVQ-LGK-PAYLIDTYQDKQDWLEGIBVGVGTAGASAPBVLNQEVIDQ 283

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 : Search time 9.98038 Seconds
(without alignments)
1346.130 Million cell updates/sec

Title: US-09-941-947a-18
Perfect score: 1623
Sequence: 1 MQIVLANPRGFCAGVDRAIE.....GIEKVFSPKELKGMQA 318

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048.5	64.6	336	4	US-09-328-352-6046 Sequence 6046, Ap
2	1028.5	63.4	349	4	US-09-199-637A-429 Sequence 429, App
3	1028.5	63.4	368	4	US-09-252-991A-21399 Sequence 21399, A
4	863	40.9	310	4	US-09-198-452A-1090 Sequence 1090, Ap
5	116.5	7.2	521	4	US-09-107-532A-6431 Sequence 6431, Ap
6	104	6.4	494	4	US-09-134-001C-5024 Sequence 5024, Ap
7	103	6.3	586	4	US-08-198-446B-19 Sequence 19, Appl
8	103	6.3	586	4	US-08-870-693-19 Sequence 19, Appl
9	101.5	6.3	540	4	US-09-134-001C-4705 Sequence 4705, Ap
10	101	6.2	547	2	US-08-467-822-35 Sequence 35, Appl
11	101	6.2	547	2	US-08-432-697-35 Sequence 35, Appl
12	101	6.2	547	3	US-08-466-248-35 Sequence 35, Appl
13	99.5	6.1	1396	1	US-08-144-121-4 Sequence 4, Appl
14	99.5	6.1	1396	2	US-08-735-893-4 Sequence 4, Appl
15	95.5	5.9	643	3	US-08-797-358B-3 Sequence 3, Appl
16	95	5.9	730	3	US-09-398-865A-2 Sequence 2, Appl
17	95	5.9	730	4	US-09-710-714-2 Sequence 2, Appl
18	94.5	5.8	516	4	US-09-134-001C-3404 Sequence 3404, Ap
19	94	5.8	496	4	US-09-202-491-7 Sequence 7, Appl
20	93	5.7	573	2	US-08-706-209-1 Sequence 1, Appl
21	93	5.7	573	3	US-08-981-787-1 Sequence 1, Appl
22	93	5.7	573	4	US-08-461-722-1 Sequence 1, Appl
23	93	5.7	573	4	US-08-336-251-1 Sequence 1, Appl
24	93	5.7	573	4	US-08-468-041-1 Sequence 1, Appl
25	93	5.7	573	4	US-08-381-861-1 Sequence 1, Appl
26	93	5.7	573	4	PCT-US94-06362-1 Sequence 1, Appl
27	93	5.7	573	5	PCT-US96-11375-1 Sequence 1, Appl

28	93	5.7	573	5	PCT-US96-11375-1 Sequence 1, Appl
29	93	5.7	981	4	US-09-252-991A-18616 Sequence 18616, A
30	91.5	5.6	298	3	US-08-961-083-24 Sequence 24, Appl
31	91.5	5.6	298	4	US-09-536-784-24 Sequence 24, Appl
32	91.5	5.6	2548	4	US-09-172-442-1 Sequence 1, Appl
33	91	5.6	335	4	US-09-934-901-6 Sequence 6, Appl
34	90.5	5.6	540	4	US-08-461-722-3 Sequence 3, Appl
35	90.5	5.6	540	4	US-08-336-251-3 Sequence 3, Appl
36	90.5	5.6	540	4	US-09-468-041-3 Sequence 3, Appl
37	90.5	5.6	541	5	PCT-US94-06362-3 Sequence 3, Appl
38	90.5	5.6	541	2	US-08-467-822-34 Sequence 34, Appl
39	90.5	5.6	541	2	US-08-447-154-19 Sequence 19, Appl
40	90.5	5.6	541	3	US-08-432-697-34 Sequence 34, Appl
41	90.5	5.6	541	3	US-08-466-248-34 Sequence 34, Appl
42	90.5	5.6	710	4	US-09-107-532A-5067 Sequence 5067, Ap
43	89.5	5.5	243	4	US-09-252-991A-18154 Sequence 18154, A
44	88.5	5.5	885	2	US-08-310-912A-2 Sequence 2, Appl
45	88.5	5.5	885	3	US-08-841-089-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-6046
; Sequence 6046, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6046
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6046

Query Match 64.6%; Score 1048.5; DB 4; Length 336;
Best Local Similarity 64.2%; Pred. No. 2.5e-98;
Matches 203; Conservative 49; Mismatches 61; Indels 3; Gaps 2;

QY	1	MQIVLANPRGFCAGVDRAIEIVDAIEAFGAPIYRREVVANRTVVDGAKGAVFIEL	60
DB	21	MEIVLANPRGFCAGVDRAIIVNRLAECPPIYRREVVANRTVVDGAKGAVFIEL	80
QY	61	SDVPVGSYLIFSANGVSKVQOBAEROLTYFDATCPITVYRMOVAHAKQGRVILIG	120
DB	81	DQVPDSDIVITSAGVSAVQOBAEROLTYFDATCPITVYRMOVAHAKQGRVILIG	140
QY	121	HAGHEVVGINGQVEKCEKTYLVEPTEDVRNKLVPNDLAVYTQTLSTMTDKVMV	180
DB	141	HGHHPVSGINGQVDPK-LKGGDIYHVEADVAALVHPRKELAFYTTLISIDTAVY	199
QY	181	DALRQPSIYKQKDDICVATQNRQDAVHDLAKISLILVGSBNSNSRLREIAVOL	240
DB	200	DALRQPSIYKQKDDICVATQNRQDAVHDLAKISLILVGSBNSNSRLREIAVOL	259
QY	241	GKPAVLITDYDLDQMDIEGIEVGVTAQAPVLYQVLDOKKAWGGETSVSENGI	300
DB	260	GKPAVLITDYDLDQMDIEGIEVGVTAQAPVLYQVLDOKKAWGGETSVSENGI	317
QY	301	EKKVFSIPKELKGM 316	
DB	318	ENITFSPKELRIHV 333	

RESULT 2
US-09-199-637A-429
; Sequence 429, Application US/09199637A

Patent No. 6355411
 GENERAL INFORMATION:
 APPLICANT: Ausubel, Frederick
 APPLICANT: Goodman, Howard W.
 APPLICANT: Rahme, Laurence G.
 APPLICANT: Mahajan-Miklos, Shalina
 APPLICANT: Tan, Man-Wah
 APPLICANT: Cao, Hui
 APPLICANT: Drenkard, Eliana
 APPLICANT: Tsongalis, John
 TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 FILE REFERENCE: 00786/361002
 CURRENT APPLICATION NUMBER: US/09/199,637A
 CURRENT FILING DATE: 1998-11-25
 PRIOR APPLICATION NUMBER: 60/066,517
 PRIOR FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 429
 LENGTH: 349
 TYPE: PR1
 ORGANISM: Pseudomonas aeruginosa
 US-09-199-637A-429

Query Match 63.4%; Score 1028.5; DB 4; Length 349;
 Best Local Similarity 63.3%; Pred. No. 2,9e-96;
 Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

QY 1 MOVLANPRGFCAGVDRALIEYDQAIIEAFGAPTYRHHVHNRTVVDGKKGAVFIEEL 60
 DB 36 MOVLANPRGFCAGVDRALIEYDQAIIEAFGAPTYRHHVHNRTVVDGKKGAVFIEEL 95
 QY 61 SDVPVGSYLIIFSAHGVSKVQDAEROLTVDPATCPLVTYKVMQVAKAKQGREVILIG 120
 DB 96 DVPDNNVIYIFSAHGVSKVQDAEROLTVDPATCPLVTYKVMQVAKAKQGREVILIG 155
 QY 121 HAHGHEVEGTMGQYKCTEGGGIYIVETPEDVBNLKVNNPNDLAVYTQTLSMDTSKVI 180
 DB 156 HEGHPEVEGTMGQYD-ASNGGAIYVDEADVAALEVRKPEALHYVTQTLSMDTSKVI 214
 QY 181 DALREQPSIKEOKKDDICATONRQDAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 240
 DB 215 DALRAKFPQIGPRKNDICATONRQDAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 274
 QY 241 GKPAVLIPTYQDLKQDMLEIGVGVGTAGASAPVYLQEVTDOLKAMGGETTSVRENSGI 300
 DB 275 GTPAYLIDGABDMQGMDFGVRIIGITAGASAPVYLQEVTDOLKAMGGETTSVRENSGI 332
 QY 301 BEKVVSIPKEIK 313
 DB 333 ENITFSMPKEIK 345

RESULT 3

US-09-252-991A-21399
 Sequence 21399, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 21399
 LENGTH: 368
 TYPE: PR1

ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-21399

Query Match 63.4%; Score 1028.5; DB 4; Length 368;
 Best Local Similarity 63.3%; Pred. No. 3.2e-96;
 Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

QY 1 MOVLANPRGFCAGVDRALIEYDQAIIEAFGAPTYRHHVHNRTVVDGKKGAVFIEEL 60
 DB 55 MOVLANPRGFCAGVDRALIEYDQAIIEAFGAPTYRHHVHNRTVVDGKKGAVFIEEL 114
 QY 61 SDVPVGSYLIIFSAHGVSKVQDAEROLTVDPATCPLVTYKVMQVAKAKQGREVILIG 120
 DB 115 DVPDNNVIYIFSAHGVSKVQDAEROLTVDPATCPLVTYKVMQVAKAKQGREVILIG 174
 QY 121 HAHGHEVEGTMGQYKCTEGGGIYIVETPEDVBNLKVNNPNDLAVYTQTLSMDTSKVI 180
 DB 175 HEGHPEVEGTMGQYD-ASNGGAIYVDEADVAALEVRKPEALHYVTQTLSMDTSKVI 233
 QY 181 DALREQPSIKEOKKDDICATONRQDAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 240
 DB 234 DALRAKFPQIGPRKNDICATONRQDAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 293
 QY 241 GKPAVLIPTYQDLKQDMLEIGVGVGTAGASAPVYLQEVTDOLKAMGGETTSVRENSGI 300
 DB 294 GTPAYLIDGABDMQGMDFGVRIIGITAGASAPVYLQEVTDOLKAMGGETTSVRENSGI 351
 QY 301 BEKVVSIPKEIK 313
 DB 352 ENITFSMPKEIK 364

RESULT 4

US-09-198-452A-1090
 Sequence 1090, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Griffiths, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment;
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 CURRENT FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 1090
 LENGTH: 310
 TYPE: PR1
 ORGANISM: Chlamydia pneumoniae
 US-09-198-452A-1090

Query Match 40.9%; Score 663; DB 4; Length 310;
 Best Local Similarity 45.0%; Pred. No. 3.7e-59;
 Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

QY 2 QIVLANPRGFCAGVDRALIEYDQAIIEAFGAPTYRHHVHNRTVVDGKKGAVFIEEL 61
 DB 3 KLILCNPRGFCAGVDRALIEYDQAIIEAFGAPTYRHHVHNRTVVDGKKGAVFIEEL 62
 QY 62 DVPVGSYLIIFSAHGVSKVQDAEROLTVDPATCPLVTYKVMQVAKAKQGREVILIG 121
 DB 63 DVPESERVYSAHGIPEVRAKAKRIIDIDATCGVTVKHSAAKIYASGGYKILIGH 122
 QY 122 AGHPEVEGTMGQYKCTEGGGIYIVETPEDVBNLKVNNPNDLAVYTQTLSMDTSKVI 181
 DB 123 KKEHVEIGIVGEVBE-----HITVEKADVBEALPSSDITPLFTITQTLSLDVQETSS 177
 QY 182 ALREQPSIKEOKKDDICATONRQDAVHDLAKISDLILVGSPPSSNSNRLREIAVOLG 241
 DB 178 ALKRYPEIITLPSSICATYATNROKALRSYLSRVYVYVGVGVNSNSNRLREIALRG 237
 QY 242 KPAVLIPTYQDLKQDMLEIGVGVGTAGASAPVYLQEVTDOLKAMGGETTSVRENSGI 301

Db 218 VPADLNNPEDIIVNHSGDIAMTAGSTPEVYQACIRL-----SSLIFGLQVE 290

QY 302 -----EYVPSIPELX 313

Db 291 NDPAVEDVAFOLPELR 308

RESULT 5
US-09-107-532A-6431
Sequence 6431, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Atinello, Pamela Denke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6431:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...521
SEQUENCE DESCRIPTION: SEQ ID NO: 6431:
US-09-107-532A-6431

Query Match 7.2%, Score 116.5; DB 4; Length 521;
Best Local Similarity 20.9%; Pred. No. 0.0035;
Matches 63; Conservative 52; Mismatches 104; Indels 83; Gaps 11;

QY 38 EVVNRRTVVDLKKQGVAFIELSDVPVGSY-----LIFSAGVSKVQQAEEQQLT 90

Db 112 EVVHTSLVBNKIKISFIABLGQPSNDLVYDLNLIIFKHENSRNLIQTTRK---- 167

QY 91 VFDAICPLVTVKQVAKQAGREVILIGHGHPVEG--TWQYKCTEGGSIYVER 148

Db 168 -----VPI-----VILEGKTGLSIQPIYSKTRKIKIAGSPFYLSS 206

QY 149 PEDVRNLKNNPNDLAVYTQTLSKTDKWK-----VDALRQPPSIKEQKDD 197

Db 207 FYDIRNKL-----LTLIVLEVSVLILSIISIGFLSSYFLKPLKVLADTDITRKQPSD 261

QY 198 ICYATONRODAVDLACISLILVSGPNSNSNRRL-----BIAVOLGKAYL 246

Db 262 IMHEPIDNDELADLAEIENKL-----DRMLYLIEQGFQVFEVSHETPTVAI 311

QY 247 IDTVQDLKQDM-LEGIEVGVTAGASAPV-----LVQEVTDOLKA-----MGQETTSV 294

Db 312 IEGLSLILNNMGDDPELLESLSASQWISRKSLVQENLDLSRAQVYVHIANETTKG 371

QY 295 RE 296

Db 372 KE 373

RESULT 6
US-09-134-001C-5024
Sequence 5024, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS FILIS REFERENCES: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5024
LENGTH: 494
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5024

Query Match 6.4%, Score 104; DB 4; Length 494;
Best Local Similarity 19.1%; Pred. No. 0.06;
Matches 64; Conservative 63; Mismatches 134; Indels 74; Gaps 12;

QY 52 KGAVFIEELSDVPVGSYLFISAGVSKVQQAEEQQLTFFDQTCPLTVKVNQVAGAK 111

Db 14 KESLTFDDVLLIPASDVLSVDVLSVKLSDKI-KNIPYISAGMDTVSSKNALIAK-AR 71

QY 112 QGSEVILIGHAG-----HP-----EYEGTQYKCTEGG 141

Db 72 QGGLGVYHKMGVEEQADDEVQKRSNGVYSNPFLLTPESVYEAALMGKTRI-----S 127

QY 142 GYLVATPEPVRNLKNNPNDLAVYTQTLSMTD-----TKWVDALRQPPSIKEQ 193

Db 128 GVPIDVNOEDRKILGILTNRDLFIEDFSIKISDVTKDNLTAPVGTTLDEABAILQKH 187

QY 194 KKQDICYATONROD--AVHDLAKISDL-----LIVSGPNSNSNRRLR-BIAVQ 239

Db 188 KIKRLPVENGRLEGLITIKDIEKYLEFPYAADENGRLLAAIGSDTSEIRAOGLVE 247

QY 240 LGRPAVLIPT-----YQDLKQDMLEGIEVGVTAGASAPVLVQEVTDOLKMG 288

Db 248 AGVDALIIDTAHSGKGVINQVGHIKETYPETIVVAGNVAATBALFAGADVAVGVI 307

QY 289 GE-----TTSVENSIGEE-KVPSIPELXKQMOA 318

Db 308 GPGSICITRVVAGVGPQITAVYDCAATEARKGKA 342

RESULT 7
US-08-198-446B-19
Sequence 19, Application US/08198446B
Patent No. 5674966
GENERAL INFORMATION:
APPLICANT: Hartwell, Leland H.
APPLICANT: Weinstein, Ted A.
APPLICANT: Plon, Sharon E.


```

APPLICANT: Groudine, Mark T.
TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/198,446B
FILING DATE: 18-Feb-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCRI17537
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-224-0779
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 586 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: Yeast MEC3 protein
US-08-198-446B-19

Query Match
Best Local Similarity 21.9%; Score 103; DB 1; Length 586;
Matches 71; Conservative 46; Mismatches 97; Indels 110; Gaps 18;

QY 34 YVRHE--VVRNRVVDGLKQKGAVFIEELSDVPVGSYLIFSA-----HGVS-----77
DB 15 YRHAAYFLRNRHTTPAQKQA-----QIQIPIENYRNFSTVAHVHGKSTLSRLLEI 69
QY 78 -----KEVQGEAE---ERQLTVFDPATCPPLVTKVMQVAKAKQREVT--LIGHA 122
DB 70 THVIDPNARKQVLDKLEVERERGITTKAQTCM-----FYKDKRTGKRYVLLHLIDTP 122
QY 123 GHPEVEGTWQ--YEKCTEGGIGIVLTPEDVRLKVNPNNDLAY-----VTQTLLS 172
DB 123 GHVDFRGVSRSYASC--GGALLVDASQGIQAVAN--FYLPFSLGKLPIVINKIDLN 179
QY 173 MDTKVMVDALRROPFSIKQKDDICATQNRQDAVHDL--AKISDLILVVGSPNSN 229
DB 180 FTDVQKQKQIYVNNF---ELPREDIIGVSRKTALNVEELLPAIDRIIPPTGRPD--- 232
QY 230 SNRLREIAVOLGRP--AYLIDTYQDLKQDWLBGIEVVGTVAGASAPVYLQEVLDQKAW 287
DB 233 -----KPRRLVDSWYD-----AYLGAVLVNIVDGF--- 260
QY 288 GGETTSVRENSGIEBKVFSIPKE 311
DB 261 -----VRKN-----DKVICAQTKR 274

```

RESULT 8
US-08-870-693-19
Sequence 19, Application US/08870693

GENERAL INFORMATION:
APPLICANT: Bartwell, Ieland H.
APPLICANT: Weinert, Ted A.
APPLICANT: Pion, Sharon E.
APPLICANT: Groudine, Mark T.

```

TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,693
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/198,446
FILING DATE: February 18, 1994
APPLICATION NUMBER: PCT/US93/04458
FILING DATE: May 12, 1993
APPLICATION NUMBER: US 07/884,426
FILING DATE: May 14, 1992
APPLICATION NUMBER: US 07/882,051
FILING DATE: May 12, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCRI10798
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-224-0779
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 586 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: Yeast MEC3 protein
US-08-870-693-19

```

Query Match
Best Local Similarity 21.9%; Score 103; DB 2; Length 586;
Matches 71; Conservative 46; Mismatches 97; Indels 110; Gaps 18;

```

QY 34 YVRHE--VVRNRVVDGLKQKGAVFIEELSDVPVGSYLIFSA-----HGVS-----77
DB 15 YRHAAYFLRNRHTTPAQKQA-----QIQIPIENYRNFSTVAHVHGKSTLSRLLEI 69
QY 78 -----KEVQGEAE---ERQLTVFDPATCPPLVTKVMQVAKAKQREVT--LIGHA 122
DB 70 THVIDPNARKQVLDKLEVERERGITTKAQTCM-----FYKDKRTGKRYVLLHLIDTP 122
QY 123 GHPEVEGTWQ--YEKCTEGGIGIVLTPEDVRLKVNPNNDLAY-----VTQTLLS 172
DB 123 GHVDFRGVSRSYASC--GGALLVDASQGIQAVAN--FYLPFSLGKLPIVINKIDLN 179
QY 173 MDTKVMVDALRROPFSIKQKDDICATQNRQDAVHDL--AKISDLILVVGSPNSN 229
DB 180 FTDVQKQKQIYVNNF---ELPREDIIGVSRKTALNVEELLPAIDRIIPPTGRPD--- 232
QY 230 SNRLREIAVOLGRP--AYLIDTYQDLKQDWLBGIEVVGTVAGASAPVYLQEVLDQKAW 287
DB 233 -----KPRRLVDSWYD-----AYLGAVLVNIVDGF--- 260
QY 288 GGETTSVRENSGIEBKVFSIPKE 311
DB 261 -----VRKN-----DKVICAQTKR 274

```

RESULT 9

US-09-134-001C-4705
 ; Sequence 4705, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4705
 ; LENGTH: 540
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4705

Query Match 6.3%; Score 101.5; DB 4; Length 540;
 Best Local Similarity 17.4%; Pred. No. 0.12;
 Matches 72; Conservative 70; Mismatches 119; Indels 153; Gaps 19;

QY 6 ANPFGCAGVDRAIRIIVDOAIEAFGAPIYVREHVHNRV-----VDGUKQ 51
 DB 111 ANPVGIRGIGIDRAVQVAIEALHISQVENKNEIAQVGAISADEIRISYSEANDKVN 170
 QY 52 KGAVIEELS-----DIPVG-----SYLISAHVSSEVOQE--ABEROLTVFD 93
 DB 171 DEVITLESNGENTLEVEVEGQPDGQYQSPYVWDSKMAELRPYLVLDKKISSFQ 230
 QY 94 ATCPITVKVMQVAKAGKQREVILIGHAGPEVEGTQYKQEGGIYIVETPEEVR 153
 DB 231 DILFLEQV-----VQASRILIVA-----DEVG-----DALT 259
 QY 154 NLKAVNP-----NDLAVYTQTLSNTDTKVMADLRQGFPSIK 191
 DB 260 NIVNRKMGQFTAVAKAPGFGDRKAMLEDAI-----LTGAQVITDIDGLE--LK 309
 QY 192 EOKKDDICATQNRQDAVHDLAKISDLIVVSPSSNSN-RLREIAVQ----- 239
 DB 310 DASIDML--GRANKVEYTKDHTVVD-----GMDENHIDARVGQIKAQIETDSSEPK 362
 QY 240 ----LGRPAVLIDTYQ-----DLK-----QDMLGJIEVY--GVTAGASAPVLAQ 278
 DB 363 KITESLGLPGVAVNQVWGSETELRKRLRIEDALNSTRAAVEGIVAGGTRALVNIY 422
 QY 279 EVIDQLKAMGETTSYR-----ENSGIEEKVVPISIPKELKKNQA 318
 DB 423 QKVSIRKAGDVEYGVNIVLAKLQAPVQIAENAGLESGIIV---EELKHA 472

RESULT 10

US-08-467-822-35
 ; Sequence 35, Application US/08467822
 ; Patent No. 5843460

; GENERAL INFORMATION:
 ; APPLICANT: Labigne, Agnes
 ; APPLICANT: Sauerbaum, Sebastien
 ; APPLICANT: Ferrero, Richard L.
 ; APPLICANT: Thibierge, Jean-Michel
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
 ; TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
 ; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fumegan, Henderson, Farabow, Garrett &
 ; ADDRESS: Dunner
 ; STREET: 1300 I Street, N.W.
 ; CITY: Washington

STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3315
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/467,822
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/447,177
 ; FILING DATE: 19-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/432,697
 ; FILING DATE: 02-MAY-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meyers, Kenneth J.
 ; REGISTRATION NUMBER: 25,146
 ; REFERENCE/DOCKET NUMBER: 03495.0137-02000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 408-4000
 ; TELEFAX: (202) 408-4400
 ; INFORMATION FOR SEQ ID NO: 35:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 547 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-467-822-35

Query Match 6.2%; Score 101; DB 2; Length 547;
 Best Local Similarity 18.2%; Pred. No. 0.14;
 Matches 72; Conservative 65; Mismatches 143; Indels 116; Gaps 16;

QY 6 ANPFGCAGVDRAIRIIVDOAIEAFGAPIYVREHVHNRV-----VVGKLR-- 50
 DB 110 ANPEIRGRGVDLAVDAVIAELKQSKQVTTPEEIAQVATISANDKEIGNISAMKVG 169
 QY 51 QKGAFT-----EELSDVP-----VGSYLIFSAHVSSEVOQE--ABEROLTVF 92
 DB 170 RKGVIYVQDKTINDELLEIEGKMFDRGYISPYFINTSKQKQCFQAAVYLLSEKXISSI 229
 QY 93 DATPELV-----TKVMQVA-----KAKQGREVILIGHAGPEVEGTMQQ 133
 DB 230 QSIIVPALEIANLVNRLKVLQVAVAPGFTLVNRLKVLQVAVAPAPGFGDRKQKQLK 289
 QY 134 YEKTEGGGIY-----LVETPEVRNLRKNNPNDLAVYTQTLSNTDTKVMAD 184
 DB 290 DMAIATGAVFGEGLTINLEDVQ-----PHDLKGVGIYIVTKDDMLLKQKDKQAIE 343
 QY 185 EOPFSIKQKDDICATQNRQDAVHDLAKISD-LIVVSPSSNSNRLR----- 234
 DB 344 KRIQEIIRQL--DVTTSEYEKELNERLATLSGVAVLKVGTSDEVEVNEKKRVYTDALN 401
 QY 235 ----EIAVOLKPAVL-----IPTYQDLKQDMLGIEIVGVGTAGASAPVLAQEV 282
 DB 402 ATPAABEGIVLGGGCLLRICFALDSLTTPANEDQKIGIEIIRTKIPR----- 451
 QY 283 QLRWGGTTSVRENSGIEEKVVPISIPKELKKNQA 318
 DB 452 ----MTIAKNAGVDSLSI-----VEKIMQS 472

RESULT 11

US-08-432-697-35
 ; Sequence 35, Application US/08432697
 ; Patent No. 6248330

QY 134 YEKTEGGGIX-----LVETPEDVRNKNPNNDLAVYVOTTLSTMTDKYMD-----ALR 184
 Db 290 DNALATGAVFGEHGTTLLEDO-----PHDLGKGVGVITKODAMLLKKGKRAQIE 343
 QY 185 EOPPSIKEOKKDDICATONRODAVHDLAKISD--LIIYVGSPTSNSNRLR----- 234
 Db 344 KRIQEIIEQL--DVTSEYKKEKLNRLAKLSDGVAVLKVGSTSDVBEVWEKKDRVTDALN 401
 QY 235 -----EIVVQKPAYL-----IDTYODLKQDLEGEVGVGTAGASAPVAVOEVID 282
 Db 402 ATRAAVEEIVGGGACALRCIPALDSLTPANEDOKIGIEIIRKTLKJPA----- 451
 QY 283 OLKAWGETTSVENGEIEKXVFSIPKELKXKMOA 318
 Db 452 -----MTAKNAGVGSGLI-----VEKIMOS 472

RESULT 13
 US-08-144-121-4
 ; Sequence 4, Application US/08144121
 ; Patent No. 5610031
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgeson, Robert E.
 ; TITLE OF INVENTION: BIX CHAIN OF LAMININ AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: BOSTON
 ; STATE: Massachusetts
 ; COUNTRY: United States
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/144,121
 ; FILING DATE: 27-OCT-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Myers, Paul L.
 ; REGISTRATION NUMBER: 35,965
 ; REFERENCE/DOCKET NUMBER: (MGP-0780.0) MGP-021
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 227-7400
 ; TELEFAX: (617) 227-5941
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1196 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; FEATURE:
 ; NAME/KEY: Domain
 ; LOCATION: 1..250
 ; FEATURE:
 ; NAME/KEY: Domain
 ; LOCATION: 251..437
 ; FEATURE:
 ; NAME/KEY: Domain
 ; LOCATION: 438..807
 ; FEATURE:
 ; NAME/KEY: Domain
 ; LOCATION: 808..840
 ; FEATURE:
 ; NAME/KEY: Domain
 ; LOCATION: 841..1196
 ; US-08-144-121-4

Query Match 6.1%; Score 99.5; DB 1; Length 1196;
 Best Local Similarity 19.8%; Fred. No. 0.71;
 Matches 52; Conservative 67; Mismatches 91; Indels 53; Gaps 13;

QY 69 LIPSAHGV--SKVQOE-----AEERQLTFDTCPLVTKVHMQVAKAQGEVILIG 120
 Db 843 LVYAHNMQAMLDQDVLNALAEVQLS-----KVSIAKLR-ADEAAQSAEDILL- 894
 QY 121 HAHPEVEGTMQYKEKTEGGGIVLVETPEDVRN- KPNPNNDLAVYVOTTLSTMTDKV 178
 Db 895 -----KTNATEKRNK-----SNEELRNILIXQIR-----FLTQSDADLDSIEA 933
 QY 179 MV-DALREOPPSIKEOKKDDICATONRODAVHDLAKISD.LIIYVGSPTSNSNRLRIA 237
 Db 934 VANEVLKXKMFSTQQLQN-----LTEDIRERVESLSQV-EVILQHSADIRAEKTLBEA 988
 QY 238 VQKGPAYLIDTYODLKQDMLGEIEVGVGTAGASAPVAVOEVIDOLKAWGETTSVREN 297
 Db 989 KRASKSATDVKVTADVWEALAEAKQVAA-----EKAIQADEBIDIGTQNLSTISSE 1043
 QY 298 SGIEKXVFSIPK--BLKXMQ 317
 Db 1044 TPASEETLFNASORISBLERVE 1066

RESULT 14
 US-08-735-893-4
 ; Sequence 4, Application US/08735893
 ; Patent No. 5914317
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgeson, Robert E.
 ; TITLE OF INVENTION: BIX CHAIN OF LAMININ AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: BOSTON
 ; STATE: Massachusetts
 ; COUNTRY: United States
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/735,893
 ; FILING DATE: 18-OCT-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION NUMBER:
 ; APPLICATION NUMBER: US 08/144,121
 ; FILING DATE: 27-OCT-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Myers, Paul L.
 ; REGISTRATION NUMBER: 35,965
 ; REFERENCE/DOCKET NUMBER: (MGP-0780.1) MGP-021DV
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 227-7400
 ; TELEFAX: (617) 227-5941
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1196 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; FEATURE:
 ; NAME/KEY: Domain
 ; LOCATION: 1..250
 ; FEATURE:

NAME/KEY: Domain
 LOCATION: 251..437
 FEATURE:
 NAME/KEY: Domain
 LOCATION: 438..807
 FEATURE:
 NAME/KEY: Domain
 LOCATION: 808..840
 FEATURE:
 NAME/KEY: Domain
 LOCATION: 841..1196
 US-08-797-358B-3

Query Match 6.1%; Score 99.5; DB 2; Length 1196;
 Best Local Similarity 19.8%; Pred. No. 0.71;

Matches 52; Conservative 67; Mismatches 91; Indels 53; Gaps 13;

QY 69 LIFSAGV---SKEVQOE---AERQUTVPDAPLVTKVHQAQKQREVLIG 120
 DB 843 LVTVAHNAWQKAMDIDQVLSALAVEQLS-----KQVSEAKLR-ADEAKQSAEDILL- 894
 QY 121 HAGHEVEGTMGQYKCTBGGGIYIVETPEVRNL--KVNPNDAVYVTTLSMTDKV 178
 DB 895 -----KVNATKEKMKR-----SNEIRNLKQIRN-----FLTQSDADLSIEA 933
 QY 179 MV-DALREQPSIKEQKDDICVATONRODAVHDLAKISDILVVGSPNSNSNRLEIA 237
 DB 934 VANETLKNMESTPQOLQN---LTEDIRRRVESLSQV-EVILQHSADADIAEMLLEEA 988
 QY 238 VQUGPAVILDTYQDLKQDMLBGIIVGVTAGASAPVYLVOEVIDQLKAMGETTSVREN 297
 DB 969 KRASSATADVKTADVMEKALEBAKQAVNA-----EKAIKQADEIDQTONMLTISSE 1043
 QY 238 SGIEKVVESIPK---ELKQMQ 117
 DB 1044 TRASETELENSQRISELRNVE 1066

RESULT 15

US-08-797-358B-3
 ; Sequence 3, Application US/08797358B
 ; Patent No. 6268478

GENERAL INFORMATION:

APPLICANT: Adams, John

TITLE OF INVENTION: INTRACELLULAR VITAMIN D BINDING PROTEIN

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/797,358B

FILING DATE: 11-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/011,491

FILING DATE: 12-FEB-1996

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-CE 3165

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
 LENGTH: 643 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 US-08-797-358B-3

Query Match 5.9%; Score 95.5; DB 3; Length 643;
 Best Local Similarity 21.9%; Pred. No. 0.67;

Matches 77; Conservative 45; Mismatches 102; Indels 127; Gaps 18;

QY 24 QATEAF-GAITYVRHEV-----HNRIVDGLQKQAVF---IEELSDVVGSLIFSAG 74
 DB 131 ETEAVYDGP--VNAVITVPAYFNDSQROATDAGATVGLAVLRINTEPTAAAI--AH 185
 QY 75 GVSKEVQOEAEERQUTVFD---ATCELVTKVHQAQKQREVLIGHAGHEVEGTM 131
 DB 186 GLDR---RGAGERNVLIFFDLGGGTFDV-----SVLSIDAGVEVAKATA 225
 QY 132 GQYKCTBGGGIYIVETPEVRNLKVN-----NPNDAVYVTTLSMTDKV 168
 DB 226 GD---THLOG-----KDPDRRLVNHFEVSEFRKRKRLSNKRALRLRTACERAKR 274
 QY 169 TILSMTDTRVVDALRQO---PSIKEQKDDICV-----ATONRODAVHDLAKIS 216
 DB 275 TLSSSTQATEIETISLFEQVVPYTSITRAKEBELCSDLFRSTLEPVEKGLADKDKXIH 334
 QY 217 DLILVGSFNSNSNRLEIRIAVQUGPAVILDTYQDLKQDMLBGIY-----VGTAG 269
 DB 335 DVYLVGSTR-----IPRVQKLIQDPFNGKBLNKSINPDEAVAYG 374
 QY 270 ASAEVAV-----QEVTDOL-----KAMGETTSVRENSGIEEK 303
 DB 375 AAVQAAVIMGDKCKRVADLLLDVAPLSLGLETRGVMTLLIQRNATIPTR 425

Search completed: January 29, 2004, 15:57:09
 Job time : 11.9804 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 : Search time 21.7073 Seconds

(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-18

Perfect score: 1623

Sequence: 1 MQVLANPRGRCAGVDRAIE.....GIEKVVSPKELKKHQA 318

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/ECT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1623	100.0	318	10	US-09-934-903-24
2	1623	100.0	318	11	US-09-941-947a-18
3	1028.5	63.4	349	11	US-09-975-719-429
4	787	48.5	338	15	US-10-156-761-10747
5	755	46.5	325	10	US-09-738-626-4643
6	743.5	45.2	329	10	US-09-712-363-271
7	733	45.2	335	10	US-09-712-363-188
8	663	40.9	310	12	US-10-289-762-1090
9	663	40.9	310	15	US-10-022-832-16
10	598.5	36.9	539	15	US-10-275-360-5
11	111	6.8	917	15	US-10-156-761-10047
12	109	6.7	578	10	US-09-712-363-282
13	109	6.7	925	12	US-10-369-493-5415
14	104.5	6.4	501	12	US-10-369-493-1052
15	104.5	6.4	1130	12	US-10-369-493-19263

16	104	6.4	869	12	US-10-369-493-19983	Sequence 19983, A
17	103	6.3	433	9	US-09-815-242-10607	Sequence 10607, A
18	103	6.3	553	12	US-10-032-585-7338	Sequence 7338, Ap
19	102	6.3	1382	10	US-09-729-653-2	Sequence 2, Appl1
20	101	6.2	391	12	US-10-369-493-20969	Sequence 20969, A
21	101	6.2	520	12	US-10-369-493-2893	Sequence 2893, Ap
22	100.5	6.2	538	12	US-10-369-493-2950	Sequence 2950, Ap
23	99.5	6.1	1765	15	US-10-037-182-8	Sequence 8, Appl1
24	99.5	6.1	1786	9	US-09-873-676-113	Sequence 113, App
25	99.5	6.1	1786	10	US-09-938-275-6	Sequence 6, Appl1
26	99.5	6.1	1786	15	US-10-037-182-6	Sequence 6, Appl1
27	97.5	6.0	607	12	US-10-369-493-19473	Sequence 19473, A
28	97.5	6.0	921	10	US-09-738-626-5659	Sequence 5659, Ap
29	97	6.0	299	12	US-10-369-493-10700	Sequence 10700, A
30	96.5	5.9	321	9	US-09-815-242-13454	Sequence 14504, A
31	96.5	5.9	1881	12	US-10-032-585-7646	Sequence 7646, Ap
32	96	5.9	952	12	US-10-108-260A-3284	Sequence 3284, Ap
33	95.5	5.9	330	12	US-10-369-493-23413	Sequence 23413, A
34	95.5	5.9	977	12	US-10-342-136-2	Sequence 2, Appl1
35	94.5	5.8	759	12	US-10-369-493-18629	Sequence 18629, A
36	94	5.8	496	15	US-10-215-224-7	Sequence 7, Appl1
37	94	5.8	496	15	US-10-263-677-9	Sequence 9, Appl1
38	94	5.8	573	10	US-09-847-6378-8	Sequence 8, Appl1
39	93.5	5.8	547	12	US-10-369-493-7917	Sequence 7917, Ap
40	93	5.7	563	12	US-10-341-434-135	Sequence 135, App
41	93	5.7	573	10	US-09-828-574-1	Sequence 1, Appl1
42	93	5.7	573	12	US-10-307-326-1	Sequence 1, Appl1
43	93	5.7	573	12	US-10-116-275-141	Sequence 141, App
44	93	5.7	573	12	US-10-367-093-12	Sequence 12, Appl1
45	93	5.7	573	15	US-10-046-649-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1

US-09-934-903-24

Sequence 24, Application US/09934903

Patent No. US20020102690A1

GENERAL INFORMATION:

APPLICANT: Koffas, Mattheos

APPLICANT: Odum, J. Martin

APPLICANT: Schenckle, Andreas J.

APPLICANT: No. US20020102690A1ton, Kelley C.

APPLICANT: Tomb, Jean-Francois

APPLICANT: Rouviere, Pierre

APPLICANT: Picataggio, Stephen

APPLICANT: Cheng, Qiong

TITLE OR INVENTION: Genes Involved in Isoprenoid Compounds Production

FILE REFERENCE: C11646 US NA

CURRENT APPLICATION NUMBER: US/09/934,903

PRIOR FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/229,907

PRIOR FILING DATE: September 1, 2001

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

SEQ ID NO 24

LENGTH: 318

TYPE: PRT

ORGANISM: Methylobionas 16a

US-09-934-903-24

Query Match 100.0%; Score 1623; DB 10; Length 318;

Best local Similarity 100.0%; Pred. No. 2.5e-150; Indels 0; Gaps 0;

Matches 318; Conservative 0; Mismatches 0;

DB

1 MQVLANPRGRCAGVDRAIEIVDGAIFAPGVVRRHVHNRITVVDGLKQKGAFTBEI 60

SDPVVSGYLLIFSANGVSKVQOEAEERQLTVFDATGCLVTKVHQAQKQKQEVLLIG 120

SDPVVSGYLLIFSANGVSKVQOEAEERQLTVFDATGCLVTKVHQAQKQKQEVLLIG 120

```

QY 121 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 180
DB 121 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 180
QY 181 DALREOFPSIKQKODICYATONRODAVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
DB 181 DALREOFPSIKQKODICYATONRODAVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
QY 241 GKPAYLIDTYODLKODMLBGIEVVGVTAGASAPVILVOEVITDOLKAWGGETTSVRENSGI 300
DB 241 GKPAYLIDTYODLKODMLBGIEVVGVTAGASAPVILVOEVITDOLKAWGGETTSVRENSGI 300
QY 301 EEKVVSIPKELKKGMOA 318
DB 301 EEKVVSIPKELKKGMOA 318

```

RESULT 2

```

US-09-941-947A-18
; Sequence 18, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: Dicosimo, Deana J.
; APPLICANT: Kofas, Matthews
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Ogden, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre B.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; US-09-941-947A-18

```

Query Match 100.0%; Score 1623; DB 11; Length 318;

Best Local Similarity 100.0%; Pred. No 2.5e-150; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKKGAVFIEEL 60
DB 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKKGAVFIEEL 60
QY 61 SDVPVGSYILFSAHGVSKVQOAEEROLTVDPATCPATKVMQVAKAKQGEVILLIG 120
DB 61 SDVPVGSYILFSAHGVSKVQOAEEROLTVDPATCPATKVMQVAKAKQGEVILLIG 120
QY 121 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 180
DB 121 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 180
QY 181 DALREOFPSIKQKODICYATONRODAVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
DB 181 DALREOFPSIKQKODICYATONRODAVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
QY 241 GKPAYLIDTYODLKODMLBGIEVVGVTAGASAPVILVOEVITDOLKAWGGETTSVRENSGI 300
DB 241 GKPAYLIDTYODLKODMLBGIEVVGVTAGASAPVILVOEVITDOLKAWGGETTSVRENSGI 300
QY 301 EEKVVSIPKELKKGMOA 318
DB 301 EEKVVSIPKELKKGMOA 318

```

DB 301 EEKVVSIPKELKKGMOA 318

RESULT 3

```

US-09-975-719-429
; Sequence 429, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Raimo, Laurence G.
; TITLE OF INVENTION: VIRULANCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-975-719-429

```

Query Match 63.4%; Score 1026.5; DB 11; Length 349;

Best Local Similarity 63.3%; Pred. No. 4.7e-92; Mismatches 63; Indels 3; Gaps 2;

```

QY 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKKGAVFIEEL 60
DB 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKKGAVFIEEL 60
QY 36 SDVPVGSYILFSAHGVSKVQOAEEROLTVDPATCPATKVMQVAKAKQGEVILLIG 120
DB 36 SDVPVGSYILFSAHGVSKVQOAEEROLTVDPATCPATKVMQVAKAKQGEVILLIG 120
QY 96 DQVDPNVIVIFSAHGVSKVQOAEEROLTVDPATCPATKVMQVAKAKQGEVILLIG 155
DB 96 DQVDPNVIVIFSAHGVSKVQOAEEROLTVDPATCPATKVMQVAKAKQGEVILLIG 155
QY 121 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 180
DB 121 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 180
QY 156 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 214
DB 156 HAHGHEVEGTMGQYEKCTGGGIIYVETPEDEVANLKVNNPNDLAYVOTLSTMTDKVM 214
QY 181 DALREOFPSIKQKODICYATONRODAVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
DB 181 DALREOFPSIKQKODICYATONRODAVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
QY 215 GKPAYLIDTYODLKODMLBGIEVVGVTAGASAPVILVOEVITDOLKAWGGETTSVRENSGI 300
DB 215 GKPAYLIDTYODLKODMLBGIEVVGVTAGASAPVILVOEVITDOLKAWGGETTSVRENSGI 300
QY 275 EEKVVSIPKELKKGMOA 318
DB 275 EEKVVSIPKELKKGMOA 318
QY 301 EEKVVSIPKELKKGMOA 318
DB 301 EEKVVSIPKELKKGMOA 318

```

RESULT 4

```

US-10-156-761-10747
; Sequence 10747, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMDA, SATOSHI
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAWAKI, YOSHIOYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089

```


1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	48.5%;	Score 787;	DB 15;	Length 338;
Best Local Similarity	48.1%;	Pred. No. 2e-68;		
Matches 153;	Conservative 71;	Mismatches 86;	Indels 8;	Gaps 3

[illegible]

```

RESULT 5
US-09-738-626-4643
? Sequence 4643, Application US/09738626
? Publication No. US20020197605A1
? GENERAL INFORMATION:
? APPLICANT: NAKAGAWA, SATOSHI
? APPLICANT: MIZOGUCHI, HIROSHI
? APPLICANT: ANDO, SEIKO
? APPLICANT: HAYASHI, MIKIRO
? APPLICANT: OCHIAI, KEIKO
? APPLICANT: YOKOI, HARUHIKO
? APPLICANT: TATEISHI, NAOKO
? APPLICANT: SENOH, AKIHIRO
? APPLICANT: IKEDA, MASATO
? APPLICANT: OZAKI, AKIO
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-125
? CURRENT APPLICATION NUMBER: US/09/738,626
? CURRENT FILING DATE: 2000-12-18
? PRIOR APPLICATION NUMBER: JP 99/377484
? PRIOR FILING DATE: 1999-12-16
? PRIOR APPLICATION NUMBER: JP 00/159162
? PRIOR FILING DATE: 2000-04-07
? PRIOR APPLICATION NUMBER: JP 00/280988
? PRIOR FILING DATE: 2000-08-03
? NUMBER OF SEQ ID NOS: 7059
? SOFTWARE: PatentIn ver. 3.0
? SEQ ID NO 4643
? LENGTH: 325
? TYPE: PRT
? ORANISM: Corynebacterium glutamicum
? US-09-738-626-4643

```

Query Match	46.5%	Score 755	DB 10	Length 325
-------------	-------	-----------	-------	------------

Best Local Similarity 48.3%; Pred. NO. 2.5e-65;
Matches 152; Conservative 67; Mismatches 86; Indels 10; Gaps 4;

Qy	2	QIYLANPREFCAGVDRALAEIYDQALEAGAGIYRHEVHNUR.TVJDGKQCAVFIELS	61
		15	15
Db	15	KILMAAPROYCAGVDRAVEYTERALAEYGAIYVRKEIYENR.YYVDLAEKGAIFVNEAS	74
		62	
Qy	62	DYVPGSYLIFSAGVSKVEYQDPAEBERQLTVDPATCPLVTYKMQVAKAAQSEVILGH	121
		75	
Db	75	EAPBGANNMFSAHGVSFWYHEBAARUKAIKALDAACPLVTYKHEVORFDRQGHILFTGH	134
		122	
Qy	122	AGHPVEGTMQYEBKCTGGGGLYVETPEUYRNLK--VNNNDLAAVYQTLTSTDPKVM	179
		135	
Db	135	BGHEEYBEGMHSVAKT-----HLVDGVAAGIATPEPLINDEPNLIYLSQTLTSTDEMEI	189
		180	
Qy	180	VDALRQPSISIEQKDDICTATONRQDAVHDLAKISDLILVGSPPNSNSNRLEPIAVQ	239
		190	
Db	190	VELKRPFPQLODPPSDIDICATONRQAVAYVARIERCELMIVGSRNSNSSVRLVEYAKQ	249
		240	
Qy	240	LEKPF--AYLIDTYQDDKQPMLEGIEVGGTAGASAPRYLVOEVIDQLAKMGAGTTSVRENS	298
		250	
Db	250	NGADNNAVLYDVARBIDPAMPFSGVELITGSSASVPBILVQGVIRLAEFGYD--DVEEVT	307
		299	
Qy	299	GIEEKVYVPSIPREK 313	
		308	
Db	308	SAAEKIYVFPALPRVLR 322	

```

1 RESULT 6
2 US-09-712-363-271
3 Sequence 271. Application US/09712263
4 Patent No. US20020164588A1
5
6 GENERAL INFORMATION:
7
8 APPLICANT: Eisenberg, David
9 APPLICANT: Roststein, Sergio H.
10 APPLICANT: Marcotte, Edward M.
11 TITLE OF INVENTION: DETERMINING THE FUNCTION OF A PROTEIN
12 TITLE OF INVENTION: INTERACTIONS OF PROTEINS
13 FILE REFERENCE: 07419-032001
14 CURRENT APPLICATION NUMBER: US/09/712,363
15
16 CURRENT FILING DATE: 2000-11-13
17 PRIORITY APPLICATION NUMBER: PCT/US00/02246
18 PRIOR FILING DATE: 2000-01-28
19
20 PRIOR APPLICATION NUMBER: 60/179,531
21 PRIOR FILING DATE: 2000-02-01
22
23 PRIOR APPLICATION NUMBER: 60/117,844
24 PRIOR FILING DATE: 1999-01-29
25
26 PRIOR APPLICATION NUMBER: 60/118,206,
27 PRIOR FILING DATE: 1999-02-01
28
29 PRIOR APPLICATION NUMBER: 60/126,593
30 PRIOR FILING DATE: 1999-03-26
31
32 PRIOR APPLICATION NUMBER: 60/134,093
33 PRIOR FILING DATE: 1999-05-14
34
35 PRIOR APPLICATION NUMBER: 60/134,092
36 PRIOR FILING DATE: 1999-05-14
37
38 PRIOR APPLICATION NUMBER: 60/165,124
39 PRIOR FILING DATE: 1999-11-12
40
41 PRIOR APPLICATION NUMBER: 60/165,086
42 PRIOR FILING DATE: 1999-11-12
43
44 SOFTWARE: FastSeq for Windows Version 4.0.0
45 SEQ ID NO: 271
46
47 LENGTH: 329
48
49 TYPE: PRT
50
51 ORGANISM: Mycobacterium tuberculosis
52 US-09-712-363-271

```

```

Query Match          45.8%; Score 743.5; DB 10; Length 329;
Best Local Similarity 47.6%; Pred. No. 3,4e-64;
Matches 150; Conservative 60; Mismatches 94; Indels 11; Gaps 3;

      3 IVANPSGFCAGDRALEIVDAIEAFGARPIYRHRHVENNRVTVDGLQKQGAVFIEELD 62
::|||:|||||:|||||:|||:::||||:|||||::|:|||||:|:|:

```

QY 3 IVLANPFCFCAGUDRALEIVDQALTEFGAPPIYVRHVEVHNRTVTDGLKKQGAVFIEELSD 62

Db 20 VLLASPRSCAGVBAEIAETVKKVLDAEGPVYVYKQIVANTVVAELDRGAVFEVDE 79
QY 63 V----PVGSYILPSAHGVSKEVQOAEERQLTVPATCPVTKVMQAKHAKQGREVIL 118
Db 80 IPDPPEPQAVVVFSAHGVSPAVFAGADERGLOVATCPVAKVAREARFPAARDIVVF 139
QY 119 IGHAGHPEVEGTMGQYKCTEGGAIYLVETPEDVRNLKVNNDLAAYTQTLSMTDKV 178
Db 140 IGHAGHPEVEGTMGQYKCTEGGAIYLVETPEDVRNLKVNNDLAAYTQTLSMTDKV 194
QY 179 MYDALREOPPSIKEQKDDICVATONRODAVHDLAKISDLILVGSFNSNSNRLREIAY 238
Db 195 VLDARAFPTLQGPSEDIQVATNRQALQSWGECDECVLVIGSCSSNSRRLVELAQ 254
QY 239 QUGKPAVILDTYQDLKQMLBGIENVGTAGASAPVAVQEVLDQKAMGETTSVRENS 298
Db 255 RSGTFAVILDPDDILEPEWLSVSTIGVAGASAPRLVGVQVLDRLGYSATL--VVERS 312
QY 299 GIEKVPFSIPKELX 313
Db 313 IATEVRFGLPKQVR 327

RESULT 7

US-09-712-363-188
; Sequence 188, Application US/09712363
; Patent No. US20020154588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032091
; CURRENT APPLICATION NUMBER: US/09/712,363
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-188

Query Match 45.2%; Score 733; DB 10; Length 335;
Best Local Similarity 47.0%; Pred. No. 3.8e-63;
Matches 147; Conservative 66; Mismatches 92; Indels 8; Gaps 3;

QY 2 QIVLANPGRFCAGVDRAIEIVDAIEAFGAPIYVRHEVNRITVVDGKQKGAVFIEELS 61
Db 27 RVLLAEPGRGVCAGVDRAIVTERLQKHGPPVYRHEIVNRRHVVDLAKGAFVVEETE 86
QY 62 DVPVGSYILPSAHGVSKEVQOAEERQLTVPATCPVTKVMQAKHAKQGREVILIGH 121
Db 87 QVPEGAIVVFSAHGVAPTVHVSASERNLQVTDATCPVTKVMQAKHAKQGREVILIGH 146

QY 122 AGHPEVEGTMGQYKCTEGGAIYLVETPEDVRNLKVNNDLAAYTQTLSMTDKVMD 181
Db 147 EGHEEVGTAGE-----APDVQLVGDVADVQVTADEDEKVMVLSQTTLSVDETWI 201
QY 182 ALREOPPSIKEQKDDICVATONRODAVHDLAKISDLILVGSFNSNSNRLREIAYOLG 241
Db 202 RLRRFRKLDPPSDDICVATONRODAVHDLAKISDLILVGSFNSNSNRLREIAYOLG 261
QY 242 -KPAVILDTYQDLKQMLBGIENVGTAGASAPVAVQEVLDQKAMGETTSVRENSGI 300
Db 262 ARAHVLVMDADIDDSAMLDGVTGVTSGASVPEVLVRGLERLAECCGYDI--VQPTT 319
QY 301 EKVFSIPKELX 313
Db 320 NETLVFALPRELR 332

RESULT 8

US-10-289-762-1090
; Sequence 1090, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1090
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-1090

Query Match 40.9%; Score 663; DB 12; Length 310;
Best Local Similarity 45.0%; Pred. No. 2.4e-56;
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

QY 2 QIVLANPGRFCAGVDRAIEIVDAIEAFGAPIYVRHEVNRITVVDGKQKGAVFIEELS 61
Db 3 KILCNCRGPGCSGVAVLQVEVALERKMGAPIYKHEIVNRRHVVALRAKGAIPVEELV 62
QY 62 DVPVGSYILPSAHGVSKEVQOAEERQLTVPATCPVTKVMQAKHAKQGREVILIGH 121
Db 63 DVEPEGERVITSAGIIPPSVRAEKARKLIDIDATCGLVTKVHSAKLYASKGKILIGH 122
QY 122 AGHPEVEGTMGQYKCTEGGAIYLVETPEDVRNLKVNNDLAAYTQTLSMTDKVMD 181
Db 123 KKEVEVIGLVGEVPE-----HITVEKVDVLEALPPSSDPLFLYIQTTLSDVDGEISS 177
QY 182 ALREOPPSIKEQKDDICVATONRODAVHDLAKISDLILVGSFNSNSNRLREIAYOLG 241
Db 178 ALKRRIPSIITLPPSSICVATTNRQALQSWGECDECVLVIGSCSSNSRRLVELARG 237
QY 242 -KPAVILDTYQDLKQMLBGIENVGTAGASAPVAVQEVLDQKAMGETTSVRENSGI 301
Db 238 VPDILINPEDIDITNIVNSGDIAMTAGASTPEDVQACIRKL-----SLIPGQVE 290
QY 302 -----EKVFSIPKELX 313
Db 291 NDIFAVEDVVFOLPRELR 308

RESULT 9

US-10-022-832-36
; Sequence 36, Application US/10022832
; Publication No. US20030059896A1
; GENERAL INFORMATION:
; APPLICANT: COUTURE, FRANCE
; APPLICANT: HAMEL, JOSEF

```

Query Match      36.9%; Score 598.5; DB 15; Length 539;
Best local Similarity 43.6%; Pred. No. 1.2e-49;
Matches 123; Conservative 55; Mismatches 99; Indels 5; Gaps 1
3 IVLANPFGCAGVDRAIEIVDQALEAGAPLYPRHRYVHNRVTVDGLKQGANFIEELSD 62

```

[illegible]

DB 880 -----MESNNIVGSPHSGSKARDVTVK 900

RESULT 12

US-09-712-363-282
 / Sequence 282, Application US/09712363
 / Patent No. US20020164588A1
 / GENERAL INFORMATION:
 / APPLICANT: Eisenberg, David
 / APPLICANT: Rotstein, Sergio H.
 / APPLICANT: Marcotte, Edward M.
 / TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 / TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
 / FILE REFERENCE: 07419-032001
 / CURRENT APPLICATION NUMBER: US/09/712,363
 / CURRENT FILING DATE: 2000-11-13
 / PRIOR APPLICATION NUMBER: PCT/US00/02246
 / PRIOR FILING DATE: 2000-01-28
 / PRIOR APPLICATION NUMBER: 60/179,531
 / PRIOR FILING DATE: 2000-02-01
 / PRIOR APPLICATION NUMBER: 60/117,844
 / PRIOR FILING DATE: 1999-01-29
 / PRIOR APPLICATION NUMBER: 60/118,206
 / PRIOR FILING DATE: 1999-02-01
 / PRIOR APPLICATION NUMBER: 60/126,593
 / PRIOR FILING DATE: 1999-03-26
 / PRIOR APPLICATION NUMBER: 60/134,093
 / PRIOR FILING DATE: 1999-05-14
 / PRIOR APPLICATION NUMBER: 60/134,092
 / PRIOR FILING DATE: 1999-05-14
 / PRIOR APPLICATION NUMBER: 60/165,124
 / PRIOR FILING DATE: 1999-11-12
 / PRIOR APPLICATION NUMBER: 60/165,086
 / PRIOR FILING DATE: 1999-11-12
 / NUMBER OF SEQ ID NOS: 292
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO: 282
 / LENGTH: 578
 / TYPE: PRT
 / ORGANISM: Mycobacterium tuberculosis
 / US-09-712-363-282

Query Match 6.7%; Score 109; DB 10; Length 578;
 Best Local Similarity 20.4%; Pred. No. 0.11;
 Matches 78; Conservative 61; Mismatches 105; Indels 138; Gaps 18;

QY 6 ANPRGFC-----AGVRAIEIVDQALZAFGAPLYREHYH 41
 DB 67 ANPCGVCESVSLAPAPGSDIVVELDASHGVDTRDLDR--AFYAPQSHRYV-- 121
 QY 42 NRTVVDGLKQKAVFIEELSDVPVGSYLLFFSAHGVSK-----EVQGEAEERQLTVD 94
 DB 122 -----FIVDEHMYVTMGFNALLKIVEPPPHLIFITAT 155
 QY 95 TCP-----LVTKVH-----MQVAKAKOGEREVILIGHAPHEVEGTMQYE 135
 DB 156 TEPEKVLPIRSTRTHYPRFLPPRTMLARICEQ--EGVVDAYPLV----- 205
 QY 136 KCTBGGGILVETPEB--VRMLKANNPVDLAYVTQT-----TLSTDTKVMQDALRQPP 188
 DB 206 -IRAGGG-----SPRUTLSVLDQLLAGAAD-THVYTRALGLLGVTDVALIDDVAV---D 254
 QY 189 SIKQCKDDICVATGNRODAVHD-----LAKISDLIVVGSFNSNS----- 230
 DB 255 ALAACDAALFGLAISVIDIGGHDPRFATDLERRODLIVOSVDAASRGVDAPEBAL 314
 QY 231 NPLREIAVLGKPAVYIIDYQDLKQDMLEGIENVGTAGASAPVLYCEVIDQL--KAW 287
 DB 315 DMKRODAIIGAT--ETRYAEVQAQGL-----GEMRGATAPRLLEVCARLLLPAS 366
 QY 288 GGETTSVRENSGIEEKVPSIP 309
 DB 367 DAEASALGVEVERIEIRLDMKIP 388

RESULT 13

US-10-369-493-5415
 / Sequence 5415, Application US/10369493
 / Publication No. US20030233675A1
 / GENERAL INFORMATION:
 / APPLICANT: Cao, Yongwei
 / APPLICANT: Hinkle, Gregory J.
 / APPLICANT: Slater, Steven C.
 / APPLICANT: Goldman, Barry S.
 / APPLICANT: Chen, Xiandeng
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 / FILE REFERENCE: 38-10(52052)B
 / CURRENT APPLICATION NUMBER: US/10/369,493
 / CURRENT FILING DATE: 2003-02-28
 / PRIOR APPLICATION NUMBER: US 60/360,039
 / PRIOR FILING DATE: 2002-02-21
 / NUMBER OF SEQ ID NOS: 47374
 / SEQ ID NO: 5415
 / LENGTH: 925
 / TYPE: PRT
 / ORGANISM: Caenorhabditis elegans
 / US-10-369-493-5415

Query Match 6.7%; Score 109; DB 12; Length 925;
 Best Local Similarity 23.1%; Pred. No. 0.23;
 Matches 77; Conservative 53; Mismatches 134; Indels 70; Gaps 17;

QY 3 IVLANPRGFCAGVRAIEIVDQALZAFGAPLYREHYHNTVVDGLKQKAVFIEELSD 62
 DB 143 MILNARKFTEEYVQAAMEIKQVVT--VPYF--TPAEELVAPRAQVAGLTVLDLIN 196
 QY 63 VPVGSYLLFFSAHGVSKVQGEAEERQLTVD--ACPLVTKVMQVAKAKOGEREVIL 118
 DB 197 DGTAAAL---SHGIFRKEIGEKGQRLMYDMGAKTATITVEKLKVEKERQPKVTVL 253
 QY 119 -IGHAGHPEVEGTMQYEKCTBGGGILVETPEDEVRLKVN--NPPDLAYVTQTLSMT 174
 DB 254 GVG-----FDRITGGIE-WTNRRLRHILIMFE--KYYKRTKVTNTRRA-MTFSKEAE 303
 QY 175 DTKVMQDALRQPPSIEQCKDDICVATGNRODAVHDLAKISDLIVVGSFNSNSRLR 234
 DB 304 RLKQVLANNEHFPQISAHBDIDAKLVTREDPNH--LISWESRFGEP-ITQALRMA 359
 QY 235 EIAVQDLKPAVYIIDYQDLKQDMLEGIENVGTAGASAPV--LYQSEVI----- 281
 DB 360 QIPV-----DDIDQFVLMG-----AGTRVPKQVEIVQKTIQTKIEIGKPLMT 400
 QY 282 DQKAMGGETTSVRENSG-----IEEKVPSI 308
 DB 401 DEAVAMGALFQNAHLSKGFVKPKNIEKVIIPV 434

RESULT 14

US-10-369-493-1052
 / Sequence 1052, Application US/10369493
 / Publication No. US20030233675A1
 / GENERAL INFORMATION:
 / APPLICANT: Cao, Yongwei
 / APPLICANT: Hinkle, Gregory J.
 / APPLICANT: Slater, Steven C.
 / APPLICANT: Goldman, Barry S.
 / APPLICANT: Chen, Xiandeng
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 / FILE REFERENCE: 38-10(52052)B
 / CURRENT APPLICATION NUMBER: US/10/369,493
 / CURRENT FILING DATE: 2003-02-28
 / PRIOR APPLICATION NUMBER: US 60/360,039
 / PRIOR FILING DATE: 2002-02-21
 / NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 1052
LENGTH: 501
TYPE: PRT
ORGANISM: Methanococcus jannaschii
US-10-369-493-1052

Query Match
Best Local Similarity 6.4%; Score 104.5; DB 12; Length 501;
Best Local Similarity 20.5%; Pred. No. 0.24;
Matches 69; Conservative 51; Mismatches 134; Indels 83; Gaps 15;

QY 14 GVDRAIEIVDAI--EAFGAPIYR-----HEVVH-RTVVDGLK 50
DB 133 GSEGAIEDIDEAIEIAEAGFPVVVYASAGGGMGVAVSKEKKEVESARNAKAF 192
QY 51 QGSAVFIEELSDVP--VGSYLIFSAHG-----VSKVQGAHEERQUTVDAICPLVTKVH 103
DB 193 GDFPIFIEKYNPNPHIEIQLGDRNGNIIHAGDRCSIORSHQKI--IEEAPSPMTB-- 249
QY 104 MQVAHAKOGREVIILGHAGHPEVSGTMQVEKTEGGGIVLVETPEDVRNKKVNPDL 163
DB 250 ---ELRRWGEALIKGKAINDSAGTV---EFLVENGFIYLE-----MNRIGVEHT- 297
QY 164 AVYQTTLSMTYKVMVDALRQPSIKEQKDDICVATQ--NFQAVHDIAKISDLIL 220
DB 298 --VTEQVTGIDLVKMIKILAGEBELTKQEDVKIRGHAIKCRINADPLNDPVPKPGKIK 355
QY 221 VVGSPPNSNSNRLRIAVQLGKPAVLIDTYQDLKQDWLEGIWGVGTAGASAPVLAQEV 280
DB 356 LYRSP-----GGPGVRIIS-----GVYGAELIIP-YIDSM 384
QY 281 IDQLKAWG--ETTSVRENSGIEKRVPSIPKELKH 315
DB 385 IAKLITYGNSREELAKMKRALREVIYIGVKTNIPIH 421

RESULT 15

US-10-369-493-19263
Sequence 19263, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xiangeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052) B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 19263
LENGTH: 1130
TYPE: PRT
ORGANISM: Myxococcus xanthus
US-10-369-493-19263

Query Match
Best Local Similarity 6.4%; Score 104.5; DB 12; Length 1130;
Best Local Similarity 16.9%; Pred. No. 0.86;
Matches 70; Conservative 64; Mismatches 124; Indels 155; Gaps 13;

QY 38 EYVNRRTVVDGLKQKGVFIEELSDV-----PVGSYLI-----FSA 73
DB 604 EYVLEADITINAMTQTLIFAQOQVTVARTVGEKLGQAQAVPGVAGTWKDLINNVNLLA 663
QY 74 HGVSKVQGAHEERQUTVDAICPLVTKVH 104
DB 664 NNLTIDGVNINAEVTTAVAKDLRSKTIYDAKGEVLELSTINTMVDQLNSPAAEVTRAK 723
QY 105 QVAKAKOGREVIILGHAGHPEVSGTMQVEKTEGGGIVLVETPEDVR----- 153

DB 724 EVTEGKLG-----GQAEVGVSGVWKDLIDNNFMAVNLTVYRGIVRVYAVAN 774
QY 154 ---NLKVNPNDLAVYTQTTLSMTDT---KVMVDALR-----EQPFSIKE 192
DB 775 GDLMQKLTMDAKGEIAMLADITINAMTQTLIFAQOQVTVARTVGEKLGQAQAEVPGVAG 834
QY 193 QKXD---DICATONRQDAVHDIKIS-----DL-----ILVGSPPNSNSNR 232
DB 835 TWKDLTNVNNILANNLTQVARNIAEVTTAVANGDLSKTIYDAKGEVLELSTINTMVDQ 894
QY 233 LREIAVOLGKPAVLIDT-----YQDLKQDWLEGIWGVGTAGASAPV----- 275
DB 895 LRAFAAVTRVAKGVETGKLGQADVDLGSVWKDLIDNNVNLAGNLTQVARNIAKVT 954
QY 276 -----LVQEVLDQLKAWGETTSVRENSGIEKV 304
DB 955 AVANGDLSQKITVSVKGEVLELKNITINTMVDQLAPASAEVTRVAKGVETGKLG 1007

Search completed: January 29, 2004, 16:21:19
Job time : 22.7073 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 9.35661 Seconds
(without alignments)
3268.453 Million cell updates/sec

Title: US-09-941-947a-18

Perfect score: 1623

Sequence: 1 MQLVLANPFGCAGVDRAIE.....GIEKKVFSIPKELKGMQA 318

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR:76:***
2: PIR:***
3: PIR:***
4: PIR:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028.5	63.4	314	2	G83076
2	1019.5	62.8	317	2	AC0059
3	1001	61.7	316	2	AD0508
4	996	61.4	316	2	JB0403
5	996	61.4	316	2	H90632
6	987	60.8	320	2	G82293
7	987	60.8	320	2	G82293
8	969	59.7	318	1	A64164
9	968	59.6	322	2	D81038
10	967	59.6	316	2	C82561
11	858	52.9	319	2	A84947
12	826.5	50.9	319	2	G87665
13	756	46.6	348	2	AH2571
14	756	46.6	348	2	F97453
15	750	46.2	346	2	AE3434
16	743.5	45.8	329	2	E70973
17	733	45.2	335	2	D70898
18	717	44.2	335	2	E87151
19	667	41.1	307	2	B71461
20	663	40.9	310	2	G72006
21	663	40.9	310	2	F86617
22	661	40.7	308	2	C81723
23	621.5	38.3	180	2	D37152
24	565	34.8	335	2	G75509
25	527	32.5	314	1	B69554
26	513	31.6	314	2	F83822
27	443.5	27.3	311	2	AC1256
28	415.5	25.6	289	2	G70449
29	369.5	22.8	277	2	F81362

30	358	22.1	642	2	H97127	fusion Penicillin
31	341	21.0	406	2	S76307	hypothetical prote
32	335.5	20.7	274	1	H64569	lys tolerance pr
33	332	20.5	275	2	A72253	lyb protein - The
34	312	19.2	402	2	AF1929	hypothetical prote
35	288	17.7	376	2	G71310	probable penicilli
36	230	14.2	452	2	T04781	hypothetical prote
37	115	7.1	929	2	T35683	flsk homolog - Str
38	112.5	6.9	434	2	B72228	pyrimidine-nucleos
39	111.5	6.9	777	1	G69773	conserved hypotet
40	110	6.8	500	2	AB1111	heat-shock protein
41	109	6.7	499	2	A11472	heat-shock protein
42	109	6.7	522	2	A33644	signal recognition
43	109	6.7	578	2	B70796	probable DNA POLYM
44	109	6.7	925	2	G88175	protein T24H7.2 [1
45	108.5	6.7	544	2	B41479	60K heat shock pro

ALIGNMENTS

RESULT 1

G83076
lyb protein PA4557 [imported] - Pseudomonas aeruginosa (strain PA01)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #ext_change 31-Dec-2000
C/Accession: G83076
R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Muzoguchi, S.D.; Warren, P.; Hickey, M.D.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A/Reference number: A82950; PMID:20437337; PMID:10984043
A/Accession: G83076
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-314 <STO>
A/Cross-references: GB:AE004869; GB:AE004091; NID:g9950793; PIDN:AA07945.1; GSPDB:GN00
A/Experimental source: strain PA01
C/Genetics:
A/Gene: lybB; PA4557
C/Superfamily: penicillin tolerance protein

Query Match 63.4%; Score 1028.5; DB 2; Length 314;
Best Local Similarity 63.3%; Pred. No. 6.8e-67;
Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

QY	1	MQLVLANPFGCAGVDRAIEYDQALBARGAPVYRHEVYVNRKTVYDGLKQGAVFIEEL	60
DB	1	MQLVLANPFGCAGVDRAIEYVNRALDVGPPVYRHEVYVNRKTVYDGLKQGAVFIEEL	60
QY	61	SDVPVGSYLIFSAHGVSKVEQAEEROLTVDATCPVATKVMQVAKAKQGREYLIG	120
DB	61	SDVPVNNVITVIFSAHGVSKVEQAEEROLTVDATCPVATKVMQVAKAKQGREYLIG	120
QY	121	HAGHPVEGCTGQYKCTEGGSIYVETPEDVRNKKVNNPNDLAVYVQTLSMTDTPKVV	180
DB	121	HAGHPVEGCTGQYKCTEGGSIYVETPEDVRNKKVNNPNDLAVYVQTLSMTDTPKVV	180
QY	181	DALREOPPSIKQKQDDICVANTONRQDAVHDIAKISDILVYVSSNSNSNRLREIAVL	240
DB	181	DALREOPPSIKQKQDDICVANTONRQDAVHDIAKISDILVYVSSNSNSNRLREIAVL	240
QY	241	GKPAVLDITDQDKQDMEGIEVVGATGASAPVAVQEVVDOLKAMGGETTSVRENSCI	300
DB	241	GKPAVLDITDQDKQDMEGIEVVGATGASAPVAVQEVVDOLKAMGGETTSVRENSCI	300
QY	301	EKKVVFSPKELK 313	
DB	301	EKKVVFSPKELK 313	
QY	298	ENITFSMPKELK 310	
DB	298	ENITFSMPKELK 310	

RESULT 2

AC0059

lytB protein [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001

C:Accession: AC0059

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Taranga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AC0059

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-317 <YUR>

A:Cross-References: GB:AL590842; PIDN:CA089334.1; PID:g15978571; GSPDB:GN00175

C:Genetics:

C:Superfamily: penicillin tolerance protein

Query Match 62.8%; Score 1019.5; DB 2; Length 317;

Best Local Similarity 63.5%; Pred. No. 3.1e-66;

Matches 200; Conservative 54; Mismatches 54; Indels 7; Gaps 4;

QY 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTYVDGKQGAFTIEL 60

DB 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTYVDGKQGAFTIEL 60

QY 61 SDVPVGSYLFSAHGSKEVQQAEEERQLT-VPDAPCPLVTKVMQVAKAGREVI 119

DB 61 SEVPDGSILIFSAHGSKEVQQAEEERQLT-VPDAPCPLVTKVMQVAKAGREVI 120

QY 120 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 179

DB 120 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 179

QY 121 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 178

DB 121 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 178

QY 180 VDAPPEPPIKQKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 239

DB 180 VDAPPEPPIKQKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 239

QY 179 IDANTRPEPKIVGPRKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 238

DB 179 IDANTRPEPKIVGPRKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 238

QY 240 LGKPAVILDTYQDKODLKGIEVGVYTAGASAPBVVGVVIOQLKAWG-GETTSTVRS 298

DB 240 LGKPAVILDTYQDKODLKGIEVGVYTAGASAPBVVGVVIOQLKAWG-GETTSTVRS 298

QY 229 MGKRAVLIDSAADIOEFNLQAKICITVTAASAPDILVQVIRLIDLANGEIT---EIS 295

DB 229 MGKRAVLIDSAADIOEFNLQAKICITVTAASAPDILVQVIRLIDLANGEIT---EIS 295

QY 299 GIEKVESIPREIK 313

DB 299 GIEKVESIPREIK 313

QY 236 GREENIVEFVPEKLR 310

DB 236 GREENIVEFVPEKLR 310

QY 236 GREENIVEFVPEKLR 310

DB 236 GREENIVEFVPEKLR 310

QY 236 GREENIVEFVPEKLR 310

DB 236 GREENIVEFVPEKLR 310

QY 236 GREENIVEFVPEKLR 310

DB 236 GREENIVEFVPEKLR 310

QY 236 GREENIVEFVPEKLR 310

DB 236 GREENIVEFVPEKLR 310

QY 236 GREENIVEFVPEKLR 310

DB 236 GREENIVEFVPEKLR 310

Best Local Similarity 62.3%; Pred. No. 6.7e-65;

Matches 195; Conservative 53; Mismatches 61; Indels 4; Gaps 2;

QY 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTYVDGKQGAFTIEL 60

DB 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTYVDGKQGAFTIEL 60

QY 61 SDVPVGSYLFSAHGSKEVQQAEEERQLT-VPDAPCPLVTKVMQVAKAGREVI 120

DB 61 SDVPVGSYLFSAHGSKEVQQAEEERQLT-VPDAPCPLVTKVMQVAKAGREVI 120

QY 61 SEVPDGSILIFSAHGSKEVQQAEEERQLT-VPDAPCPLVTKVMQVAKAGREVI 120

DB 61 SEVPDGSILIFSAHGSKEVQQAEEERQLT-VPDAPCPLVTKVMQVAKAGREVI 120

QY 121 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 180

DB 121 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 180

QY 121 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 178

DB 121 GHAGPEVETMGVYKCTEGGGLYVETPEDVKNKNPNDAVYVOTTTSMPTKYM 178

QY 181 DALREPPSIKQKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 240

DB 181 DALREPPSIKQKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 240

QY 179 DALREPPSIKQKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 238

DB 179 DALREPPSIKQKODICATONRQDAVDLAKISDLILVGSNNSSNRLREIAVQ 238

QY 241 GKPAVILDTYQDKODLKGIEVGVYTAGASAPBVVGVVIOQLKAWG-GETTSTVRS 300

DB 241 GKPAVILDTYQDKODLKGIEVGVYTAGASAPBVVGVVIOQLKAWG-GETTSTVRS 300

QY 239 GRAPFIDDAADIOEFNLQAKICITVTAASAPDILVQVIRLIDLANGEIT---GR 296

DB 239 GRAPFIDDAADIOEFNLQAKICITVTAASAPDILVQVIRLIDLANGEIT---GR 296

QY 301 EKKVFSIPREIK 313

DB 301 EKKVFSIPREIK 313

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 297 ENIVEFVPEKLR 309

DB 297 ENIVEFVPEKLR 309

QY 61 SDVPGVSYLIFSAHGVSKVQQAEROLTVFADATCPVTVKVMQVAKHAKQREVLIG 120
 Db 61 SEVDDGAILIFSAHGVSKVQQAEROLTVFADATCPVTVKVMQVAKHAKQREVLIG 120
 QY 121 HAGHPEVGTMGQYKCEGGGILVETPEDVRLKVNPNDAVYQTTLSTMTDKVMV 180
 Db 121 HAGHPEVGTMGQYKCEGGGILVETPEDVRLKVNPNDAVYQTTLSTMTDKVMV 180
 QY 181 DALREOPPSIKKQKDDICVATONRODAVHDLAKISDLILVVGSPNSNSNRRLAEIAYOL 240
 Db 179 DALRKRPFKVGPBKDDICVATONRODAVHDLAKISDLILVVGSPNSNSNRRLAEIAYOL 238
 QY 241 GKRAVYLDITVODLKQMLBEGIEVVGTVAGASAPVTVQVETDOLKAMGETTSVRENSGI 300
 Db 239 GKRAFLIDDAKDIOEHWKVKCVGTAGASAPDILVQNVARLQOLGGGEALPLE--GR 296
 QY 301 EKKVPSIPKELK 313
 Db 297 EENIVEVPEKELR 309

RESULT 5

LybB protein [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: H90632
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhsara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 6, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genod
 A:Reference number: A96629; MUID:21156231; PMID:11258796
 A:Accession: H90632
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-316 <HAY>
 A:Cross-references: GB:BA000007; PTDN:BAE33455.1; PTD:G13359488; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: EC50032
 C:Superfamily: penicillin tolerance protein

Query Match 61.4%; Score 996; DB 2; Length 316;
 Best Local Similarity 61.3%; Pred. No. 1.5e-64;
 Matches 192; Conservative 59; Mismatches 58; Indels 4; Gaps 2;
 QY 1 MQLIANPRGFCAGVDRAIEIVDAIEAFGAPVYRHEVYHNRVTVDGKQKGAFFIEEL 60
 Db 1 MQLIANPRGFCAGVDRAIEIVDAIEAFGAPVYRHEVYHNRVTVDGKQKGAFFIEEL 60
 QY 61 SDVPGVSYLIFSAHGVSKVQQAEROLTVFADATCPVTVKVMQVAKHAKQREVLIG 120
 Db 61 SEVDDGAILIFSAHGVSKVQQAEROLTVFADATCPVTVKVMQVAKHAKQREVLIG 120
 QY 121 HAGHPEVGTMGQYKCEGGGILVETPEDVRLKVNPNDAVYQTTLSTMTDKVMV 180
 Db 121 HAGHPEVGTMGQYKCEGGGILVETPEDVRLKVNPNDAVYQTTLSTMTDKVMV 180
 QY 181 DALREOPPSIKKQKDDICVATONRODAVHDLAKISDLILVVGSPNSNSNRRLAEIAYOL 240
 Db 179 DALRKRPFKVGPBKDDICVATONRODAVHDLAKISDLILVVGSPNSNSNRRLAEIAYOL 238
 QY 241 GKRAVYLDITVODLKQMLBEGIEVVGTVAGASAPVTVQVETDOLKAMGETTSVRENSGI 300
 Db 239 GKRAFLIDDAKDIOEHWKVKCVGTAGASAPDILVQNVARLQOLGGGEALPLE--GR 296
 QY 301 EKKVPSIPKELK 313
 Db 297 EENIVEVPEKELR 309

RESULT 6

85483
 hypothetical protein lybB [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: G85483
 R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grochbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousie, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85483
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-316 <STO>
 A:Cross-references: GB:AE005174; NID:G12512712; PTDN:AA054331.1; GSPDB:GN00145; UKGP:200
 A:Experimental source: strain O157:H7, substrain EDL93
 C:Genetics:
 A:Gene: lybB
 C:Superfamily: penicillin tolerance protein

Query Match 61.4%; Score 996; DB 2; Length 316;
 Best Local Similarity 61.3%; Pred. No. 1.5e-64;
 Matches 192; Conservative 59; Mismatches 58; Indels 4; Gaps 2;
 QY 1 MQLIANPRGFCAGVDRAIEIVDAIEAFGAPVYRHEVYHNRVTVDGKQKGAFFIEEL 60
 Db 1 MQLIANPRGFCAGVDRAIEIVDAIEAFGAPVYRHEVYHNRVTVDGKQKGAFFIEEL 60
 QY 61 SDVPGVSYLIFSAHGVSKVQQAEROLTVFADATCPVTVKVMQVAKHAKQREVLIG 120
 Db 61 SEVDDGAILIFSAHGVSKVQQAEROLTVFADATCPVTVKVMQVAKHAKQREVLIG 120
 QY 121 HAGHPEVGTMGQYKCEGGGILVETPEDVRLKVNPNDAVYQTTLSTMTDKVMV 180
 Db 121 HAGHPEVGTMGQYKCEGGGILVETPEDVRLKVNPNDAVYQTTLSTMTDKVMV 180
 QY 181 DALREOPPSIKKQKDDICVATONRODAVHDLAKISDLILVVGSPNSNSNRRLAEIAYOL 240
 Db 179 DALRKRPFKVGPBKDDICVATONRODAVHDLAKISDLILVVGSPNSNSNRRLAEIAYOL 238
 QY 241 GKRAVYLDITVODLKQMLBEGIEVVGTVAGASAPVTVQVETDOLKAMGETTSVRENSGI 300
 Db 239 GKRAFLIDDAKDIOEHWKVKCVGTAGASAPDILVQNVARLQOLGGGEALPLE--GR 296
 QY 301 EKKVPSIPKELK 313
 Db 297 EENIVEVPEKELR 309

RESULT 7

882293
 lybB protein VC0685 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C:Species: Vibrio cholerae
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: G82293
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, K.D.; Vamathavan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: AB2035; MUID:20406833; PMID:10952301
 A:Accession: G82293
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-320 <HEI>
 A:Cross-references: GB:AE004154; GB:AE003852; NID:G9655115; PTDN:AAF93850.1; GSPDB:GN001;
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC0685
 A:Map position: 1
 C:Superfamily: penicillin tolerance protein

Query Match

60.8%; Score 987; DB 2; Length 320;

Best Local Similarity 61.9%; Pred. No. 7e-64;
Matches 195; Conservative 56; Mismatches 58; Indels 6; Gaps 3;

```
QY 1 MQVLNPRGFCAGVDRAIEIVDAIEAFGAPLYVEHEVHNRTVVDGLKQKGAFTIEBL 60
DB 5 MKILLNPRGFCAGVDRAISIVERALELYOPPIYVEHEVHNRTVVDGLKQKGAFTIEBL 64
QY 61 SDVPVGSYLIFSAHGVSKVQGEAEEROLTVDPATCPILTKVMQVAKAKQGREVILIG 120
DB 65 NEVPDDNIVIFSAHGVSKVQGEAEEROLTVDPATCPILTKVMQVAKAKQGREVILIG 124
QY 121 HAHGPEVEGTMGQY--ASDTGMYLVEKPEVDVSLQAKVDPENLHAYGQTTLSVDETMD 178
DB 125 HAHGPEVEGTMGQY--ASDTGMYLVEKPEVDVSLQAKVDPENLHAYGQTTLSVDETMD 182
QY 179 MVDALEQPFISKEQKDDICVATONRODAVHDLAKISDLILVSGSPNSNRLREIAY 238
DB 183 VIDELERRVFEKIQPKRDDICVATONRODAVHDLAKISDLILVSGSPNSNRLREIAY 242
QY 239 QLGKPAVILDTYQDLKQDMLEGIEVVGVTAGASAPFVLQVQEVIDQLKMGGETTSVENS 298
DB 243 KLGTPGLIDCDPDIDPEMFVDAQILGVTAGASAPFVLQVQEVIDQLKMGGETTSVENS 300
QY 299 GIEKAVVFSIPKELK 313
DB 301 GREENMFEVPEKELQ 315
```

RESULT 8

A64164
lybB protein homolog H11007 - Haemophilus influenzae (strain Rd KX20)

C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A:Accession: A64164
R:Leischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kitchess, E.F.; Kervayage, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhman, J.L.; Geoghegan, N.S.K. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A. Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: A64164
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-318 <TIGR>
A:Cross-references: GB:L42023; TIGR:H11007
C:Genetics:
A:Start codon: GTG
C:Superfamily: penicillin tolerance protein

Query Match 59.7%; Score 969; DB 1; Length 318;
Best Local Similarity 58.8%; Pred. No. 1.4e-62;
Matches 184; Conservative 61; Mismatches 64; Indels 4; Gaps 2;

```
QY 1 MQVLNPRGFCAGVDRAIEIVDAIEAFGAPLYVEHEVHNRTVVDGLKQKGAFTIEBL 60
DB 5 MKILLNPRGFCAGVDRAISIVERALELYOPPIYVEHEVHNRTVVDGLKQKGAFTIEBL 64
QY 61 SDVPVGSYLIFSAHGVSKVQGEAEEROLTVDPATCPILTKVMQVAKAKQGREVILIG 120
DB 65 NEVPDDNIVIFSAHGVSKVQGEAEEROLTVDPATCPILTKVMQVAKAKQGREVILIG 124
QY 121 HAHGPEVEGTMGQY--ASDTGMYLVEKPEVDVSLQAKVDPENLHAYGQTTLSVDETMD 178
DB 125 HAHGPEVEGTMGQY--ASDTGMYLVEKPEVDVSLQAKVDPENLHAYGQTTLSVDETMD 182
QY 181 DAIREQFPIKQKDDICVATONRODAVHDLAKISDLILVSGSPNSNRLREIAY 240
DB 183 AALKEKYPALIQGHKDDICVATONRODAVHDLAKISDLILVSGSPNSNRLREIAY 242
QY 241 GKRAVILDTYQDLKQDMLEGIEVVGVTAGASAPFVLQVQEVIDQLKMGGETTSVENS 300
DB 243 GVKSQLIDEPADIQDMFVNDVKTGITAGASAPFVLQVQEVIDQLKMGGETTSVENS 300
```

QY 301 EKVVSIPKELK 313
DB 301 BGNMFEVPEKELR 313

RESULT 9

DB1038

lybB protein NME1831 [imported] - Neisseria meningitidis (strain MC58 serogroup B, strain C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001

R:Retzlaff, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Sisen, J.A.; Hickey, B.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Yi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizzi, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, A. Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: DB1038
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <TIGR>
A:Cross-references: GB:A8002532; GB:A8002098; NID:97227078; PIDN:AAF42166.1; PID:9722708
A:Experimental source: serogroup B, strain MC58
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, J.; Holtz, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, N. Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491. A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: DB1982
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:97379120; PIDN:CA863914.1; PID:9737935
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: lybB; MMB1831; NMA0624
C:Superfamily: penicillin tolerance protein

Query Match 59.6%; Score 968; DB 2; Length 322;
Best Local Similarity 61.5%; Pred. No. 1.7e-62;
Matches 192; Conservative 48; Mismatches 66; Indels 6; Gaps 2;

```
QY 3 IVLANPRGFCAGVDRAIEIVDAIEAFGAPLYVEHEVHNRTVVDGLKQKGAFTIEBL 62
DB 6 IILANPRGFCAGVDRAISIVERALEBFAGPIYVEHEVHNRTVVDGLKQKGAFTIEBL 65
QY 63 VPVGSYLIFSAHGVSKVQGEAEEROLTVDPATCPILTKVMQVAKAKQGREVILIG 122
DB 66 VPVGSYLIFSAHGVSKVQGEAEEROLTVDPATCPILTKVMQVAKAKQGREVILIG 125
QY 123 GHPEVEGTMGQY--GKMLVEIVGDVAKLEVRNPKDILAYSGQTTLSVDETMD 182
DB 126 GHPEVEGTMGQY--GKMLVEIVGDVAKLEVRNPKDILAYSGQTTLSVDETMD 181
QY 183 LREQPFPIKQKDDICVATONRODAVHDLAKISDLILVSGSPNSNRLREIAY 242
DB 182 LREQPFPIKQKDDICVATONRODAVHDLAKISDLILVSGSPNSNRLREIAY 241
QY 243 PAVILDTYQDLKQDMLEGIEVVGVTAGASAPFVLQVQEVIDQLKMGGETTSVENS 302
DB 242 DAYVNDNAGVILQRAAMPBPKNNVGTAGASAPFVLQVQEVIDQLKMGGETTSVENS 299
QY 303 KVVFSIPKELK 314
DB 300 SIIVLPKELR 311
```

RESULT 10

C82561

drug tolerance protein XE2416 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa

QY 181 DALREDFPSIKQKDDICATONRQDAVHDLAKISDLILVWSPNSNSNRLEIAVQL 240
 DB 187 ALREREPGIAAPKXEDICATTRRQDAVHDLAKISDLILVWSPNSNSNRLEIAVQL 246
 QY 241 G-KPAILIDYQDLKQDMLEGIEVGTAGASAPVLYQVYDQKMGCTTSVRENSG 239
 DB 247 GARARLIDIASGIDMTWFDGIRVGLTAGASAPEDLVQVYDAISA--RFDITVELVE 304
 QY 300 IEEKVFSIPKEL 312
 DB 305 AERTITFKLRLL 317

RESULT 13

penicillin tolerance protein (imported) - Agrobacterium tumefaciens (strain C58, Dupont)
 A:Accession: AH2671
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AH2671
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, T.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; Mclell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.M.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:21608550; PMID:11743193

A:Accession: AH2671
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-348 <KUR>
 A:Cross-references: GB:AB008688; PIDN:AA141790.1; PID:G17739144; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:

A:Gene: lycB
 A:Map position: circular chromosome
 C:Superfamily: penicillin tolerance protein

Query Match 46.6%; Score 756; DB 2; Length 348;
 Best Local Similarity 50.5%; Pred. No. 3.7e-47;

Matches 161; Conservative 54; Mismatches 90; Indels 14; Gaps 6;

QY 1 MOVLAMPFGCAGVDRAIEVDQAIKAFGAPIVRHEVHNRTVVDGLKQKAVFEEL 60
 DB 10 LTRLCGRPGFCAGVDRAIQIVLALKAQAPVYRHEIYHNRYVEGLAKGALFEEL 69
 QY 61 SDVPEV--GSTLIFSAHGSKEVQAEERQLTVPDTCPLTKVMQVAKAKQREVI 117
 DB 70 HEIPAEHREQPVFSAHGVKSPVEDAQARNLFYLDATCPLVSKVHKQARHQLGRHV 129
 QY 118 LIGHAGHPEVGTQMEQYEKCTEGGIGIYLVETPEDEVANLKVNPNIDLAAYTQTLISMTDK 177
 DB 130 LIGHAGHPEVIGTMQLPE---GTVSLVETVEDAGVYEVDRNIGFVYQTLISVDDTA 185
 QY 178 VMVDALREDFPSIKQKDDICATONRQDAVHDLAKISDLILVWSPNSNSNRLEIA 217
 DB 186 GVIARLQERFPALQAPADSIQVATNRQDAVHDLAKISDLILVWSPNSNSNRLEIA 245
 QY 238 VQUG-KPAILIDYQDLKQDMLEGIEVGTAGASAPVLYQVYDQKMGCTTSV 294
 DB 246 LRAAKHSHVLPQASFI--DMNEIGDIRTVGLSAGASAPVLYDEIIKAFKRFDTL 303
 QY 295 RENSGIEKVFSPKEL 313
 DB 304 AVT--VETHEFLVNRRL 320

RESULT 14

lyb protein (AB006203) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: F97453

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
 A.; Liu, F.; Mollan, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappes, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:21608551; PMID:11743194
 A:Accession: F97453
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-348 <KUR>
 A:Cross-references: GB:AB007869; PIDN:AAK6583.1; PID:G15155749; GSPDB:GN00186
 C:Genetics:
 A:Gene: AGR_C1414
 A:Map position: circular chromosome
 C:Superfamily: penicillin tolerance protein

Query Match 46.6%; Score 756; DB 2; Length 348;
 Best Local Similarity 50.5%; Pred. No. 3.7e-47;

Matches 161; Conservative 54; Mismatches 90; Indels 14; Gaps 6;

QY 1 MOVLAMPFGCAGVDRAIEVDQAIKAFGAPIVRHEVHNRTVVDGLKQKAVFEEL 60
 DB 10 LTRLCGRPGFCAGVDRAIQIVLALKAQAPVYRHEIYHNRYVEGLAKGALFEEL 69
 QY 61 SDVPEV--GSTLIFSAHGSKEVQAEERQLTVPDTCPLTKVMQVAKAKQREVI 117
 DB 70 HEIPAEHREQPVFSAHGVKSPVEDAQARNLFYLDATCPLVSKVHKQARHQLGRHV 129
 QY 118 LIGHAGHPEVGTQMEQYEKCTEGGIGIYLVETPEDEVANLKVNPNIDLAAYTQTLISMTDK 177
 DB 130 LIGHAGHPEVIGTMQLPE---GTVSLVETVEDAGVYEVDRNIGFVYQTLISVDDTA 185
 QY 178 VMVDALREDFPSIKQKDDICATONRQDAVHDLAKISDLILVWSPNSNSNRLEIA 217
 DB 186 GVIARLQERFPALQAPADSIQVATNRQDAVHDLAKISDLILVWSPNSNSNRLEIA 245
 QY 238 VQUG-KPAILIDYQDLKQDMLEGIEVGTAGASAPVLYQVYDQKMGCTTSV 294
 DB 246 LRAAKHSHVLPQASFI--DMNEIGDIRTVGLSAGASAPVLYDEIIKAFKRFDTL 303
 QY 295 RENSGIEKVFSPKEL 313
 DB 304 AVT--VETHEFLVNRRL 320

RESULT 15

lyb protein [imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002

C:Accession: AB3434
 R:Belvecchio, V.G.; Kaprat, V.; Redkar, R.U.; Patre, G.; Muier, C.; Joe, T.; Ivanova,
 .; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AB3434
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-346 <KUR>
 A:Cross-references: GB:AB008917; PIDN:AA152640.1; PID:G17983462; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:

A:Gene: BME11459
 A:Map position: I
 C:Superfamily: penicillin tolerance protein

Query Match 46.2%; Score 750; DB 2; Length 346;
 Best Local Similarity 47.7%; Pred. No. 1e-46;

Matches 154; Conservative 59; Mismatches 88; Indels 22; Gaps 5;

QY 1 MOVLAMPFGCAGVDRAIEVDQAIKAFGAPIVRHEVHNRTVVDGLKQKAVFEEL 60
 DB 8 LTRLCGRPGFCAGVDRAIQIVLALKAQAPVYRHEIYHNRYVEGLAKGALFEEL 67

QY 61 SDVPV---GSYLLFSAGVSKVEVOQSAEROLTFVDFATCPVYTKVHMVAKHAKOGREVT 117
 Db 68 DEIPAHBNQPVVFSAGVSKVADAEAKNLFYLDATCPVSKVHKQAMRHQRLGRHVI 127
 QY 118 LIGHAGHPEVEGTMGQYEKCTEGGCIYLVETPEDVBNLKVNPNNDLAYVTQTTLISMTDX 177
 Db 128 LISHSGHPEVIGTMGQ---LPDGAVTLLIETVEDAHTCHFDDEDNLGFVQTTLISVDDTA 183
 QY 178 VMVDALREOPPSIKEQXODDICYATORODAVHDLAKISJLLVGSPPNSNSNRLEITA 237
 Db 184 GIITKELOAKFPNLAAPAESICVATTNRQDVAAPAGCDLFLVGA PNSSNSKRLVEVA 243
 QY 238 VQLG-KPAYLIDYQDLKQDMLBGEIEVVGATAGASAPVWQEVTDOLKANG-----GE 290
 Db 244 EKAGARMSMLVQRAEDIEWQIGDISVGLSAGASAPEIIVDEIIDAFAKAPDVXIELAE 303
 QY 291 TTSVRENSGIEEKVPSIPEKELK 313
 Db 304 TTVETEN-----FLVNRREIR 318

Search completed: January 29, 2004, 15:55:34
 Job time : 10.3566 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 5.98823 seconds
(without alignments)
2497.314 Million cell updates/sec

Title: US-09-941-947A-18

Perfect score: 1623

Sequence: 1 MOIYLANPRGFCAGVDRALE.....GIEKVFESIPRELKKNQQA 318

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1064.5	65.6	316	1 ISPH_ACTICA	Q9xj0 acinetobact
2	1028.5	63.4	314	1 ISPH_PSEAE	Q9hmt7 pseudomonas
3	1019.5	62.8	317	1 ISPH_YARPE	P58680 yersinia pe
4	1001	61.7	316	1 ISPH_SALT1	P58678 saltanella
5	996	61.4	316	1 ISPH_ECOLI	P22655 escherichia
6	996	61.4	316	1 ISPH_SALTY	P58679 salmonella
7	989.5	61.0	321	1 ISPH_VIBPA	Q87877 vibrio para
8	989	60.9	320	1 ISPH_VIBBT	Q8dco vibrio vuln
9	987	60.8	316	1 ISPH_VIBCH	Q9xk44 vibrio chol
10	980	60.4	316	1 ISPH_XANAC	Q8p17 xanthomonas
11	970	59.8	316	1 ISPH_XANCP	Q8p094 xanthomonas
12	969	59.7	314	1 ISPH_HAEIN	P44976 haemophilus
13	969	59.7	314	1 ISPH_PASMT	P57960 pasteurella
14	968	59.6	322	1 ISPH_NEIMA	Q91c39 neisseria m
15	967	59.6	322	1 ISPH_XYLFA	Q9p89 xyloella fas
16	949.5	58.5	324	1 ISPH_RALSO	P58677 raietonia s
17	886	53.4	315	1 ISPH_MIGBR	Q8d12r wiggleswort
18	858	52.9	319	1 ISPH_BUCAP	P57247 buchnera ap
19	826.5	50.9	319	1 ISPH_CAUCR	Q9a345 callobacter
20	824.5	50.8	313	1 ISPH_BUCAP	Q84924 buchnera ap
21	813.5	50.1	313	1 ISPH_BURPS	Q9f10 burkholderi
22	791	48.7	338	1 ISPH_STRCO	Q9dml streptomyce
23	756	46.6	348	1 ISPH_AGRTS	P58673 agrobacteri
24	755	46.5	325	1 ISPH_CORGL	Q8hmt2 corynobacte
25	752.5	46.4	350	1 ISPH_RHIME	Q92t92 rhizobium m
26	750	46.2	346	1 ISPH_BRUME	Q84yfr1 bruceella me
27	743.5	45.6	336	1 ISPH_RHIMO	O5409 mycobacteri
28	740	45.6	336	1 ISPH_RHIMO	O5409 mycobacteri
29	733	45.2	335	1 ISPH_MYCCE	O5458 mycobacteri
30	717	44.2	332	1 ISPH_MYCCE	Q84781 mycobacteri
31	667	41.1	307	1 ISPH_CHLTP	O84867 chlamydia c
32	663	40.9	310	1 ISPH_CHLTP	O926p2 chlamydia p
33	661	40.7	308	1 ISPH_CHLTP	O9p159 chlamydia m

34	621.5	38.3	180	1 ISPH_PSEFL	P21864 pseudomonas
35	565	34.8	335	1 ISPH_DEIRA	Q9xg0 deinococcus
36	527	32.5	314	1 ISPH_BACSD	P54473 bacillus su
37	513	31.6	314	1 ISPH_BACSD	Q9xk37 bacillus ha
38	443.5	27.3	331	1 ISPH_LISMO	P58676 listeria mo
39	433	26.7	288	1 ISPH_THETN	Q8476 thermosmar
40	415.5	25.6	289	1 ISPH_AQUAE	O6725 aquifex ae
41	389.5	24.0	282	1 ISPH_CLOBE	P58675 clostridium
42	369.5	22.8	277	1 ISPH_CAMBE	P94644 campylobact
43	341	21.0	379	1 ISPH_SYNT3	Q55643 synchocyst
44	335.5	20.7	274	1 ISPH_HLUPY	O25160 helicobacte
45	332	20.5	275	1 ISPH_THEMA	Q9x1f7 thermotoga

ALIGNMENTS

RESULT 1	ID	ISPH_ACTICA	STANDARD	PRT	316 AA.
AC	Q9xj0:				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	ISPH protein.				
DE	ISPH OR LYTB.				
OS	Acinetobacter calcoaceticus.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Moraxellaceae; Acinetobacter.				
OX	NCBI_TaxID=471;				
RM	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BD413 / ADP1;				
RX	MEDLINE=20224825; PubMed=10763755;				
RA	Herzberg C., Friedrich A., Averhoff B.;				
RT	"Comb, a novel competence gene required for natural transformation of				
RT	Acinetobacter sp. BD413: identification, characterization and analysis				
RT	of growth phase-dependent regulation.";				
RL	Arch. Microbiol. 173:220-228(2000).				
CC	-1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate				
CC	into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate				
CC	(DMAPP) (by similarity).				
CC	-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh				
CC	(last) step.				
CC	-1- SIMILARITY: BELONGS TO THE ISPH FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL: AF027189; AAC55804.1; -				
DR	HAMAP: MF_00191; -; 1.				
DR	InterPro: IPR003451; LYTB.				
DR	Pfam: PF02401; LYTB; 1.				
DR	TIGRFAMs: TIGR00216; ISPH_LYTB; 1.				
KM	ISOPRENE biosynthesis.				
SC	SEQUENCE 316 AA; 34943 MW; 29F071CE2B9A762 CRC64;				
Qy	Query Match	65.6%;	Score 1064.5;	DB 1;	Length 316;
Qy	Best Local Similarity	65.2%;	Pred. No. 2.1e-66;		
Db	Matches 204;	Conservative 48;	Mismatches 58;	Indels 3;	Gaps 2;
Qy	1 MOIYLANPRGFCAGVDRALEIVDAIPARFPIVREHVNNRVVVGKQKQGVFIETL 60				
Db	1 MEIYLANPRGFCAGVDRALEIVDAIPARFPIVREHVNNRVVVGKQKQGVFIETL 60				
Qy	61 SDPEVGSYLLFSAAGVKEVQQAEEERQLTFPATCELVTKVHNVQAKAKQGREVILIG 120				
Db	61 DQVVDSDIVIFSAAGVKEVQQAEEERQLTFPATCELVTKVHNVQAKAKQGREVILIG 120				

QY 121 HAGPEVEGTMGQYKCTEGGGIYLVETPEVDYANLKNPNNDLAYVYOTTLSTMTDKRMV 180
 DB 122 HEGHPEVEGTMGQYK-SKGGHLYVEDEADYALANVEPELAFYOTTLSTMTDKRMV 179
 QY 181 DALREOPFSIKQKDDICATQNRQDAVADLAKISDLILVGSPPNSNSNRLEIAVOL 240
 DB 180 DALRTFPOIQGPRKNDICATQNRQDAVADLAKISDLILVGSPPNSNSNRLEIAVOL 239
 QY 241 GKPAYLIDTYODLQDMLEGEIVGVYTAGASAPVLYOEVYDOLKAMGETTSYRENSGI 300
 DB 240 GKPAYLIDTYODLQDMLEGEIVGVYTAGASAPVLYOEVYDOLKAMGETTSYRENSGI 297
 QY 301 BEKVFSIPKELK 313
 DB 298 EENITFSIPKELR 310

RESULT 2

ISPH_PSEAE STANDARD; PRT; 314 AA.
 AC 09HVM7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ISPH protein.
 GN ISPH OR LYTB OR PA4557.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RA MEDLINE=20437337; PubMed=10984043;
 RA Steyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yan Y.,
 RA Brody L.L., Collier S.N., Folger K.R., Kas A., Laidig K., Lim R.W.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter U., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).

CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 CC (DMAPP) (by similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 CC (last) step.

CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

CC EMBL: AE004863; AAC07945.1; --
 CC FIR: G83076; G83076.
 CC HAMAP: MF_00191; -- 1.
 CC InterPro: IPR003451; IYB.
 CC Pfam: PF02401; LYTB.1.
 CC TIGRFAMs: TIGR00216; isph_lytb.1.
 CC KW isoprene biosynthesis; Complete proteome.
 CC SEQUENCE 314 AA; 34762 MW; 483E07B6B663EB6 CRC64;

Query Match 63.4%; Score 1028.5; DB 1; Length 314;
 Best Local Similarity 63.3%; Pred. No. 7.5e-66;
 Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

QY 1 MQIVLANRGACAGDRAIEIVDAIEARGAFIYRHEVHNRTYVVDLQKQGVFIIEEL 60

DB 1 MQIVLANRGACAGDRAIEIVDAIEARGAFIYRHEVHNRTYVVDLQKQGVFIIEEL 60
 QY 61 SDVPGSLTISAGVSEKVEQDEARQITVFDATCPVYTVHMQVAKHAGSEVILIG 120
 DB 61 DQVPMNVVIPSAGVSGAVKKEARGRLKVFDAICPLVTVHMEVARSDEGECVILIG 120
 QY 121 HAGPEVEGTMGQYKCTEGGGIYLVETPEVDYANLKNPNNDLAYVYOTTLSTMTDKRMV 180
 DB 121 HEGHPEVEGTMGQYD-ASNGAILYVEDEADYALAEVAKPEALHYVOTTLSTMTDKRMV 179
 QY 241 GKPAYLIDTYODLQDMLEGEIVGVYTAGASAPVLYOEVYDOLKAMGETTSYRENSGI 300
 DB 240 GKPAYLIDTYODLQDMLEGEIVGVYTAGASAPVLYOEVYDOLKAMGETTSYRENSGI 297
 QY 301 BEKVFSIPKELK 313
 DB 298 EENITFSIPKELR 310

RESULT 3

ISPH_YERPE STANDARD; PRT; 317 AA.
 AC P56880;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ISPH protein.
 GN ISPH OR LYTB OR YPO0477 OR Y3697.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Orientalis;
 RA MEDLINE=21470413; PubMed=11586360;
 RA Partridge J.B., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Parkhill J., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Terriza A.M.,
 RA Chillingworth T., Cronin A., Davies R.W., Davis P., Dougan G.,
 RA Feltham T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
 RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).

CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 CC (DMAPP) (by similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 CC (last) step.

CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AJ14143; CAC8934.1; -
CC or send an email to license@isb-sib.ch.
DR EMBL: AE013973; AAM87245.1; ALT_INIT.
DR PIR: AC0059; AC0059.
DR HAMAP: MF_00191; -; 1.
DR InterPro: IPR003451; LYTB.
DR Pfam: PF02401; LYTB; 1.
DR TIGRFAMs: TIGR00216; isph_lytb; 1.
KM isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 317 AA; 34708 MW; C40B953DCBA12168 CRC64;

Query Match 62.8%; Score 1019.5; DB 1; Length 317;
Best Local Similarity 63.5%; Pred. No. 3.3e-65;
Matches 200; Conservative 54; Mismatches 54; Indels 7; Gaps 4;

QY 1 MOVLANPRGFCAGVDRAIETVDALFAFGAPYVHREYVHNTVVDLKKOKAVPIEEL 60
1 MOVLANPRGFCAGVDRAIETVDALFAFGAPYVHREYVHNTVVDLKKOKAVPIEEL 60
DB 61 SDVPVGSYLIFSAHGVSKVQOAEEROLTVDPATCPLVTXVMQVAKAKGSEVITL 119
61 SEVPDGSILIFSAHGVSKVQOAEEROLTVDPATCPLVTXVMQVAKAKGSEVITL 120
QY 120 GHAGHPEVSGTMOQYKCTEGGIVYVTEDEVNTKVNPNDAVYTQTLLSMITKMW 179
121 GHAGHPEVSGTMOQYKCTEGGIVYVTEDEVNTKVNPNDAVYTQTLLSMITKMW 178
QY 180 VDLAROPFPIKQKODICVATONRODAVHDLAKISDLIVVGSPPNSNSNLAETAVQ 239
179 IDLNTREFPIVGRDDICVATONRODAVHDLAKISDLIVVGSPPNSNSNLAETAVQ 238
QY 240 IGRPAVLDITYODLKODMLEGIEVGVGTAGASAPVLYOEVLDOLKAWG-GFTTSVENS 298
239 MGRPAVLDITYODLKODMLEGIEVGVGTAGASAPVLYOEVLDOLKAWG-GFTTSVENS 295
QY 299 GIEKVFSPKELK 313
296 GREENIVFEVPEKLR 310

RESULT 4

ISPH_SALT1 STANDARD; PRT; 316 AA.

AC 558678;
DB 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ISPH protein.
CN ISPH OR LYTB OR STY0058 OR T0051.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
NCBI TaxID=601;
[1]
RP SEQUENCE FROM N.A.

STRAIN=CT18;
MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mangall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Hargreaves A., Hien T.T., Holtroyd S., Jagers K.,
RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).

RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22511367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Barland V., Kodoyianni V., Schwartz D.C., Blatter F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).

CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(B)-butenyl 4-diphosphate
CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
CC (DMAPP) (By similarity).
CC -1- PATHWAY: Monomevalonate terpenoid biosynthesis pathway; seventh
CC (last) step.
CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: A627265; CAD01204.1; -
DR EMBL: AE016834; AAO67784.1; -
DR HAMAP: MF_00191; -; 1.
DR InterPro: IPR003451; LYTB.
DR Pfam: PF02401; LYTB; 1.
DR TIGRFAMs: TIGR00216; isph_lytb; 1.
KM isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 316 AA; 34513 MW; 9536031F11F6DC42 CRC64;

Query Match 61.7%; Score 1001; DB 1; Length 316;
Best Local Similarity 62.3%; Pred. No. 6.7e-64;
Matches 195; Conservative 53; Mismatches 61; Indels 4; Gaps 2;

QY 1 MOVLANPRGFCAGVDRAIETVDALFAFGAPYVHREYVHNTVVDLKKOKAVPIEEL 60
1 MOVLANPRGFCAGVDRAIETVDALFAFGAPYVHREYVHNTVVDLKKOKAVPIEEL 60
DB 61 SDVPVGSYLIFSAHGVSKVQOAEEROLTVDPATCPLVTXVMQVAKAKGSEVITL 120
61 SEVPDGSILIFSAHGVSKVQOAEEROLTVDPATCPLVTXVMQVAKAKGSEVITL 120
QY 121 HAHGHEVSGTMOQYKCTEGGIVYVTEDEVNTKVNPNDAVYTQTLLSMITKMW 180
122 HAHGHEVSGTMOQYKCTEGGIVYVTEDEVNTKVNPNDAVYTQTLLSMITKMW 178
QY 181 DALROFPPIKQKODICVATONRODAVHDLAKISDLIVVGSPPNSNSNLAETAVQ 240
179 DALROFPPIKQKODICVATONRODAVHDLAKISDLIVVGSPPNSNSNLAETAVQ 238
QY 241 GKPAYLIDITYODLKODMLEGIEVGVGTAGASAPVLYOEVLDOLKAWG-GFTTSVENS 300
239 GKPAYLIDITYODLKODMLEGIEVGVGTAGASAPVLYOEVLDOLKAWG-GFTTSVENS 296
QY 301 BEKVFSPKELK 313
297 ENIVFEVPEKLR 309.

RESULT 5

ISPH_ECOLI STANDARD; PRT; 316 AA.

AC P22565;
DB 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ISPH protein.
CN ISPH OR LYTB OR B0029 OR C0033 OR Z0034 OR EGS0032.
OS Escherichia coli O6 and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562, 217992, 83334;
 RN [1] SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=21819424; PubMed=11818558;
 RA Reichelt F., Hecht S., Gaertner K., Adam P., Krieger C., Anslinger S.,
 RA Arigoni D., Bacher A., Eisenreich W.,
 RT "Studies on the nonmevalonate terpenoid biosynthetic pathway: metabolic
 RT role of Isph (lytB) protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:1158-1163(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=91187617; PubMed=2011499;
 RA Bouvier J., Stragier P.,
 RT "Nucleotide sequence of the isp-dapB interval in *Escherichia coli*.";
 RL Nucleic Acids Res. 19:180-180(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RA Isono K., Mizobuchi K., Nakata A.,
 RT "Systematic sequencing of the *Escherichia coli* genome: analysis of
 RT the 0-2.4 min region.";
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor U., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06.H / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasdo K.A., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.,
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157.H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor U., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.,
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157.H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kubota S., Shida T., Hattori M., Shinagawa H.,
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genetic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [8]
 RP IDENTIFICATION.
 RX MEDLINE=93163053; PubMed=8432714;
 RA Gustafson C.E., Kaul S., Ishiguro E.E.,

RT "Identification of the *Escherichia coli* lytB gene, which is involved
 RT in penicillin tolerance and control of the stringent response.";
 RL J. Bacteriol. 175:1203-1205(1993).
 RN [9]
 RP PATHWAY.
 RC STRAIN=K12 / ATCC 23716 / DSM 498;
 RX MEDLINE=2111595; PubMed=11418107;
 RA Altincicek B., Kollas A.-K., Eberl M., Wiesner J., Sanderbrand S.,
 RA Hintz M., Beck E., Jomaa H.,
 RT "lytB, a novel gene of the 2-C-methyl-D-erythritol 4-phosphate pathway
 RT of isoprenoid biosynthesis in *Escherichia coli*.";
 RL FEBS Lett. 499:37-40(2001).
 RN [10]
 RP PATHWAY.
 RX MEDLINE=21574179; PubMed=11717301;
 RA McAteer S., Coulson A., McLeenan N., Masters M.,
 RT "The lytB gene of *Escherichia coli* is essential and specifies a
 RT product needed for isoprenoid biosynthesis.";
 RL J. Bacteriol. 183:7403-7407(2001).
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(3)-butenyl 4-diphosphate
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 CC (DMAPP). Is also involved in penicillin tolerance and control of
 CC the stringent response. Seems to directly or indirectly interact
 CC with rda to maintain it in an inactive form during normal growth.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 CC (last) step.
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AY062212; AAL38655.1; -
 DR EMBL, X54945; CAA38707.1; -
 DR EMBL, D10483; BAB96598.1; -
 DR EMBL, AB000113; AAC73140.1; -
 DR EMBL, AB016755; AAN78533.1; ALT_INIT.
 DR EMBL, AB005179; AAG54331.1; -
 DR EMBL, AP002550; BAB33455.1; -
 DR PIR, G85483; G85483.
 DR PIR, G85483; G85483.
 DR PIR, H90632; H90632.
 DR PIR, J60403; J60403.
 DR Ecogen; Bcl1081; isph.
 DR HAMAP; MF 00191; -; 1.
 DR InterPro; IPR003451; lytB.
 DR Pfam; PF02401; lytB; 1
 DR TIGRfam; TIGR00216; isph_lytB; 1
 KM Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 316 AA; 34774 MW; 0E7B378BD49AD771 CRC64;
 Query Match 61.4%; Score 996; DB 1; Length 316;
 Best Local Similarity 61.3%; Pred. NO. 1.5e-63;
 Matches 192; Conservativity 59; Mismatches 56; Indels 4; Gaps 2;
 QY 1 MOITANPRGFCAGDRAIETDQAIETAFGAPITVRHEVNNRTTYDGLKQGANFISRL 60
 DB 1 MOILLANPRGFCAGDRAIETDQAIETAFGAPITVRHEVNNRTTYDGLKQGANFISRL 60
 QY 61 SVVPAGSYLIFSAHVSKEVQOEAEERQLTVFDATCPVTFKVMQVAKAKQGREVILIG 120
 DB 61 SVVPAGSYLIFSAHVSKEVQOEAEERQLTVFDATCPVTFKVMQVAKAKQGREVILIG 120
 QY 121 FAGHPEVEGTGQYKCTEGGSIYVETPEDEVKLNKVNPNDAVYVOTTSMTDKVMV 180
 DB 121 FAGHPEVEGTGQYKCTEGGSIYVETPEDEVKLNKVNPNDAVYVOTTSMTDKVMV 180
 QY 122 FAGHPEVEGTGQYKCTEGGSIYVETPEDEVKLNKVNPNDAVYVOTTSMTDKVMV 180
 DB 122 FAGHPEVEGTGQYKCTEGGSIYVETPEDEVKLNKVNPNDAVYVOTTSMTDKVMV 180
 QY 161 DALRROPFSIKQKKDDICVAFQNRQDAVHDAKTSDLILVVGSPNSNSRLREIAVOE 240
 DB 161 DALRROPFSIKQKKDDICVAFQNRQDAVHDAKTSDLILVVGSPNSNSRLREIAVOE 240
 QY 179 DALRRKRPCTVGRKDDICVAFQNRQDAVHDAKTSDLILVVGSPNSNSRLREIAVOE 240
 DB 179 DALRRKRPCTVGRKDDICVAFQNRQDAVHDAKTSDLILVVGSPNSNSRLREIAVOE 240

QY 241 GKPAYLIDTYODLKQDMLGIEVGTAGASAPDEVLYQVVIDOLKMGGETTSVRENSGI 300
 DB 239 GKRAFLIDDAKDQGEVVKVCKCVCVAGASAPDILVQVNVARLQGLGGEALPLE--GR 296
 QY 301 EEKVVSIPKELK 313
 DB 297 EENIVEVEPKELR 309

RESULT 6

ISPH_SALTY STANDARD; PRT; 316 AA.
 AC PS8679;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE IspH Protein.
 GN IspH OR LYTB OR STM0049.
 OS Salmomella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmomella.
 OX NCBI_TaxID=602;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=L72 / SCS0412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McDaniel M., Sanderson K.E., Spiech J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du S., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grevell N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Neam M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmomella enterica serovar Typhimurium
 L72.";
 RL Nature 413:852-856(2001).

CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 (DMAPP) (By similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 (last) step.

CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AF008695; AL19013.1;
 DR StyGene; SG7777; ISPH.
 DR HAMAP; MF_00191; -; 1.
 DR InterPro; IPR003451; LYCB.
 DR Pfam; PF02401; LYTB; 1.
 DR TIGRFAMs; TIGR00216; isph LYTB; 1.
 KW Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 316 AA; 34527 MW; 953151A411FB522 CRC64;

Query Match 61.4%; Score 996; DB 1; Length 316;
 Best Local Similarity 62.0%; Pred. No. 1.5e-63;
 Matches 194; Conservative 54; Mismatches 61; Indels 4; Gaps 2;

QY 1 MOIYLANPGRGFCAGVDALIEYDQALFAFGAPITYRHHVHNRTVYDGLKQKGAFFIEL 60
 DB 1 MOIYLANPGRGFCAGVDALIEYDQALFAFGAPITYRHHVHNRTVYDGLKQKGAFFIEL 60
 QY 61 SDVPVGSYLLIFSAHGVSKVQOEAEEROLTVFATCPLYTKVHMQVAKHAKGSEVILIG 120
 DB 61 SEVPDGAIIIFSAHGVSAVNRHAKSRDLTVFATCPLYTKVHMEVAAASRGESILIG 120
 QY 121 HAGHEVEGTMGQYKCTBGGSIIYVETPEDVRLKVNPNDAVYVOTTLISMTDTKYV 180
 DB 121 HAGHEVEGTMGQYKCTBGGSIIYVETPEDVRLKVNPNDAVYVOTTLISMTDTKYV 180

DB 121 HAGHEVEGTMGQYS--NPEGWLYVESPEDVRLVNVKNGKLSFMTQTLISVDTSDVI 178
 QY 181 DALREPFSEIKQKQDCTCATQNRDAVHDLAKISDLILVGSPPNSNSNRLEIAVOL 240
 DB 179 DALRKFPXIVGPRKDDICVATTNRQEAVALAEOQVAVLVGVSXKSSNSNRLEIAQGM 238
 QY 241 GKPAYLIDTYODLKQDMLGIEVGTAGASAPDEVLYQVVIDOLKMGGETTSVRENSGI 300
 DB 239 GKRAFLIDDAKDQGEVVKVCKCVCVAGASAPDILVQVNVARLQGLGGEALPLE--GR 296
 QY 301 EEKVVSIPKELK 313
 DB 297 EENIVEVEPKELR 309

RESULT 7

ISPH_VIBPA STANDARD; PRT; 321 AA.
 AC Q87587;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE IspH protein.
 GN IspH OR LYTB OR VP0537.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nishimura M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 distinct from that of V. cholerae.";
 RL Lancet 361:743-749(2003).

CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 (DMAPP) (By similarity).

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 (last) step.

CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AP005074; BACS8800.1;
 DR HAMAP; MF_00191; -; 1.
 DR Isoprene biosynthesis; Complete proteome.
 KW SEQUENCE 321 AA; 35721 MW; DEB4ABA3970798 CRC64;

Query Match 61.0%; Score 989.5; DB 1; Length 321;
 Best Local Similarity 61.9%; Pred. No. 4.5e-63;
 Matches 195; Conservative 54; Mismatches 61; Indels 5; Gaps 3;

QY 1 MOIYLANPGRGFCAGVDALIEYDQALFAFGAPITYRHHVHNRTVYDGLKQKGAFFIEL 60
 DB 5 MKIILANPGRGFCAGVDALIEYDQALFAFGAPITYRHHVHNRTVYDGLKQKGAFFIEL 64
 QY 61 SDVPVGSYLLIFSAHGVSKVQOEAEEROLTVFATCPLYTKVHMQVAKHAKGSEVILIG 120
 DB 65 HEVPDNNIVIFSAHGVSAVNRHAKSRDLTVFATCPLYTKVHMEVAAASRGESILIG 124
 QY 121 HAGHEVEGTMGQYKCTBGGSIIYVETPEDVRLKVNPNDAVYVOTTLISMTDTKYV 178
 DB 125 HAGHEVEGTMGQYS--SEYGMVLYETPADVRLKAIIVDPSDLHVSQTLISVDETAD 182

QY 179 MDALREOPPSIKKQKODICATONRODAVHDLAKISDLILVWSPNSNSNRLREIAY 238
 Db 183 VIDELRRPFPEIÖGRKODICATONRODAVHDLAKISDLILVWSPNSNSNRLREIAY 242
 QY 239 QUGKPAVILDTYODLKODMLEGIEVVGATAGASAEVILVOEVIDOLKAMGETTSVRENS 298
 Db 243 KIGTPAVILDTYODLKODMLEGIEVVGATAGASAEVILVOEVIDOLKAMGETTSVRENS 301
 QY 299 GIEKRVFSPKELK 313
 Db 302 GREENMFEVPEKELQ 316

RESULT 8

ISPH_VIBU STANDARD; PRT; 320 AA.
 ID ISPH_VIBU
 AC QGDET0;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE ISPH protein.
 GN ISPH OR LYTB OR VVI0504.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 NCBI_TaxID=672;
 RN SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.B.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 (DMAPP) (By similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 (last) step.
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: AB016798; AAC09023.1;
 DR HAMAP: MF_00191; 1.
 DR InterPro: IPR003451; LYCB.
 DR Pfam: PF02401; LYTB; 1.
 DR TIGRFAMs: TIGR00216; isph lytb; 1.
 DR Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 320 AA; 35575 MW; 09F8454ACDA131D8 CRC64;

Query Match 60.8%; Score 989; DB 1; Length 320;
 Best Local Similarity 62.2%; Pred. No. 4, 8e-63;

Matches 196; Conservative 52; Mismatches 61; Indels 6; Gaps 3;

QY 1 MOVIANRGFCAGVRAIEIVDAIEAFGAPIYVREHVNRVTYDGLKQKAVFIEL 60
 Db 5 MKLILANRGFCAGVRAISIVRAIEIVDAIEAFGAPIYVREHVNRVTYDGLKQKAVFIEL 64
 QY 61 SDVPVGSYLIFSAGVSKVQQAERQQLTVFDPATCPVTYKVMQVAKAKQGREVILIG 120
 Db 65 HEVPDNIIVIFSAGVSKVQQAERQQLTVFDPATCPVTYKVMQVAKAKQGREVILIG 124
 QY 121 HACHPEVSGMGQYKCTEGGQYIVETPEDVANK--VNNPDILAYVYQTLISMTDTAV 178
 Db 125 HACHPEVSGMGQYKCTEGGQYIVETPEDVANK--VNNPDILAYVYQTLISMTDTAV 182

QY 179 MDALREOPPSIKKQKODICATONRODAVHDLAKISDLILVWSPNSNSNRLREIAY 238
 Db 183 VIDELRRPFPEIÖGRKODICATONRODAVHDLAKISDLILVWSPNSNSNRLREIAY 242
 QY 239 QUGKPAVILDTYODLKODMLEGIEVVGATAGASAEVILVOEVIDOLKAMGETTSVRENS 298
 Db 243 KIGTPAVILDTYODLKODMLEGIEVVGATAGASAEVILVOEVIDOLKAMGETTSVRENS 300
 QY 299 GIEKRVFSPKELK 313
 Db 301 GREENMFEVPEKELQ 315

RESULT 9

ISPH_VIBU STANDARD; PRT; 316 AA.
 ID ISPH_VIBU
 AC QGKUA4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ISPH protein.
 GN ISPH OR LYTB OR VC0685.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 NCBI_TaxID=666;
 RN SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heldelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.R., Peterson J.D., Unayen L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Rasmussen M.D., Vamathevan J., Baas S., Qin H., Drago I., Sellers P.,
 RA McDonald L., Ulfersback T., Fleischmann R.D., Mierman M.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RL Nature 406:477-483 (2000).
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 (DMAPP) (By similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 (last) step.
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: AE004154; AAF93850.1; ALT_INIT.
 DR HAMAP: MF_00191; 1.
 DR InterPro: IPR003451; LYCB.
 DR Pfam: PF02401; LYTB; 1.
 DR TIGRFAMs: TIGR00216; isph lytb; 1.
 DR Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 316 AA; 35156 MW; BB309DBA08C679E CRC64;

Query Match 60.8%; Score 987; DB 1; Length 316;
 Best Local Similarity 61.9%; Pred. No. 6, 6e-63;

Matches 195; Conservative 56; Mismatches 58; Indels 6; Gaps 3;

QY 1 MOVIANRGFCAGVRAIEIVDAIEAFGAPIYVREHVNRVTYDGLKQKAVFIEL 60
 Db 1 MKLILANRGFCAGVRAISIVRAIEIVDAIEAFGAPIYVREHVNRVTYDGLKQKAVFIEL 60
 QY 61 SDVPVGSYLIFSAGVSKVQQAERQQLTVFDPATCPVTYKVMQVAKAKQGREVILIG 120

```

Db 61 NEVDDDDIVIFSAAGVSGAARQAEKERSLTVFDTATCLTVKMHREVARARSKHEVLLIG 120
QY 121 HAGPEVEGTMGQYKTEGGGILVETPEDEVNRL--KNNPNDAVYVQTTLSTMTDKV 178
Db 121 HAGPEVEGTMGQYKTEGGGILVETPEDEVNRL--KNNPNDAVYVQTTLSTMTDKV 178
Db 179 MDALREQFSPSIEKOKKDDICATONRQDAVHDLAKISDLIVVGSPPSSNSNRLEIRIAV 238
QY 179 VIDELRAVFPKIQGPRKDDICATONRQDAVHDLAKISDLIVVGSPPSSNSNRLEIRIAV 238
Db 239 QLGKPAVLIITVQDLKODMLEGIEVGVITGASAPVLYVQEVITDQLKAMGGETTSVRENS 298
QY 239 KLGTPGLIIPCOPDIDPEFVMDQLIGVTAGASAPVLYVQEVITDQLKAMGGETTSVRENS 298
Db 299 GIEKVPFSPISPKELK 313
QY 299 GIEKVPFSPISPKELK 311
Db 297 GREENMFEVFPKELQ 311

```

RESULT 10

```

ISPH_XANAC STANDARD; PRT; 316 AA.
AC 08PBT;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
ISPH protein.
ISPH OR LYTB OR XAC1256.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=92829;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Qaaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
Camarote G., Cammaro F., Cardoso J., Chambergro F., Chapina L.P.,
Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
-1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
(DMAPP) (by similarity).
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
(last) step.
-1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC EMBL: A011755; AAMB6128.1; -
CC HAMAP: MF_00191; -1.
CC InterPro: IPR003451; LYTB.

```

```

DR Pfam: PF02401; LYTB; 1.
DR TIGRNAME: TIGR00216; isph_lytb; 1.
KM isoprene biosynthesis; Complete proteome.
SQ SEQUNCE 316 AA; 34690 KM; C93125D6A53D4DA CRC64;

```

Query Match 60.4%; Score 980; DB 1; Length 316;

Best Local Similarity 61.3%; Pred. No. 2.1e-62;

Matches 192; Conservative 50; Mismatches 69; Indels 2; Gaps 1;

```

QY 1 MVLILANPGFCAGVDRAIEIVDQAIKAFGAPLYTHREVENRTVYDGLKQGANVIEEL 60
Db 1 MVLILANPGFCAGVDRAIEIVDQAIKAFGAPLYTHREVENRTVYDGLKQGANVIEEL 60
QY 61 SDVPVGSYVIFPSHAGVSKVQGBAEEROLTFPDACTPLTKVMQYAKAKQREVLLIG 120
Db 61 SDVPVGSYVIFPSHAGVSKVQGBAEEROLTFPDACTPLTKVMQYAKAKQREVLLIG 120
QY 61 DEVPDPAIVIFSAHGVSGAARQAEKERSLTVFDTATCLTVKMHREVARARSKHEVLLIG 120
Db 61 DEVPDPAIVIFSAHGVSGAARQAEKERSLTVFDTATCLTVKMHREVARARSKHEVLLIG 120
QY 121 HAGPEVEGTMGQYKTEGGGILVETPEDEVNRL--KNNPNDAVYVQTTLSTMTDKV 180
Db 121 HAGPEVEGTMGQYKTEGGGILVETPEDEVNRL--KNNPNDAVYVQTTLSTMTDKV 180
QY 121 HAGPEVEGTMGQYKTEGGGILVETPEDEVNRL--KNNPNDAVYVQTTLSTMTDKV 180
Db 121 HAGPEVEGTMGQYKTEGGGILVETPEDEVNRL--KNNPNDAVYVQTTLSTMTDKV 180
QY 181 DALREQFSPSIEKOKKDDICATONRQDAVHDLAKISDLIVVGSPPSSNSNRLEIRIAV 240
Db 181 DALREQFSPSIEKOKKDDICATONRQDAVHDLAKISDLIVVGSPPSSNSNRLEIRIAV 240
QY 181 EALRAVFPKIQGPRKDDICATONRQDAVHDLAKISDLIVVGSPPSSNSNRLEIRIAV 240
Db 181 EALRAVFPKIQGPRKDDICATONRQDAVHDLAKISDLIVVGSPPSSNSNRLEIRIAV 240
QY 241 GXPAYLIDTVQDLKODMLEGIEVGVITGASAPVLYVQEVITDQLKAMGGETTSVRENSGI 300
Db 241 GXPAYLIDTVQDLKODMLEGIEVGVITGASAPVLYVQEVITDQLKAMGGETTSVRENSGI 300
QY 241 GXPAYLIDTVQDLKODMLEGIEVGVITGASAPVLYVQEVITDQLKAMGGETTSVRENSGI 300
Db 241 GXPAYLIDTVQDLKODMLEGIEVGVITGASAPVLYVQEVITDQLKAMGGETTSVRENSGI 300
QY 301 EKVVPFSPISPKELK 313
Db 301 EKVVPFSPISPKELK 311

```

RESULT 11

```

ISPH_XANAC STANDARD; PRT; 316 AA.
AC 08PBT;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
ISPH protein.
ISPH OR LYTB OR XCC1157.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=340;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Qaaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
Camarote G., Cammaro F., Cardoso J., Chambergro F., Chapina L.P.,
Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
-1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
(DMAPP) (by similarity).
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh

```

(last) step.
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: A6012314; AAM40456.1; -
 DR HAMAP: MF_00191; -
 DR InterPro: IPR003451; LytB.
 DR Pfam: PF02401; LytB; 1.
 DR TIGRFAMs: TIGR00216; isph_lytB; 1.
 DR Isoprene biosynthesis; Complete proteome.
 KM
 SQ SEQUENCE 316 AA; 34641 MW; 6BFA272A4CA1EAC CRC64;

Query Match 59.8%; Score 970; DB 1; Length 316;
 Best local similarity 61.3%; Pred. No. 1.1e-61;
 Matches 192; Conservative 48; Mismatches 71; Indels 2; Gaps 1;

QY 1 MOVLANPRGFCAGVDRAIEIVDAQIAEFGAPIYVHEVHNRTVVDGLKQGAVFIEEL 60
 DB 1 MDVLANPRGFCAGVDRAIEIVDAQIAEFGAPIYVHEVHNRTVVDGLKQGAVFIEEL 60
 QY 61 SDVPGSYLIFSAGVSEVQAEEROLTFPDATCPVTKVMQVAKAKQREYVILIG 120
 DB 61 DEVPDATVIFSAGVSEVQAEEROLTFPDATCPVTKVMQVAKAKQREYVILIG 120
 QY 121 HAHGPEVEGTMQYEKCTEGGGIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 DB 121 HAHGPEVEGTMQYSESGERGQIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 QY 121 HAHGPEVEGTMQYSESGERGQIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 DB 121 HAHGPEVEGTMQYSESGERGQIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 QY 181 DALRQPSIKQKQKDDICATONRODAVNDLAKISDLILVYSGSPSSNSNRRLRIAYOL 240
 DB 181 DALRQPSIKQKQKDDICATONRODAVNDLAKISDLILVYSGSPSSNSNRRLRIAYOL 240
 QY 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 DB 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 QY 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 DB 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 QY 301 EEKVVSIPKEIK 313
 DB 299 PESWFEALPEIKR 311

RESULT 12
 ISPH_HABIN STANDARD; PRT; 314 AA.
 ID ISPH_HABIN STANDARD; PRT; 314 AA.
 AC P44976;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ISPH protein.
 GN ISPH OR LYTB OR H11007.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 NCBI_TaxID=727;
 RX MEDLINE=95350630; PubMed=7542800;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KM20 / ATCC 51907;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 Scott J.D., Shiley R., Fitzhugh W., Fields C.A., Gocayne J.D.,
 Weidman J.F., Phillips C.A., Spriggs T., Hedlorn E., Cotton M.D.,
 Uterback T.R., Hanna M.C., Nguyen D.T., Sandek D.M., Brandon R.C.,
 Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.K., Smith H.O.,

RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd".
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 CC (DMAPP) (by similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 CC (last) step.
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U32781; AAC22668.1; -
 DR TIGR: H11007; -
 DR HAMAP: MF_00191; -
 DR InterPro: IPR003451; LytB.
 DR Pfam: PF02401; LytB; 1.
 DR TIGRFAMs: TIGR00216; isph_lytB; 1.
 DR Isoprene biosynthesis; Complete proteome.
 KM
 SQ SEQUENCE 314 AA; 34468 MW; 44860D110E9F9EBC CRC64;

Query Match 59.7%; Score 969; DB 1; Length 314;
 Best local similarity 58.8%; Pred. No. 1.2e-61;
 Matches 184; Conservative 61; Mismatches 64; Indels 4; Gaps 2;

QY 1 MOVLANPRGFCAGVDRAIEIVDAQIAEFGAPIYVHEVHNRTVVDGLKQGAVFIEEL 60
 DB 1 MDVLANPRGFCAGVDRAIEIVDAQIAEFGAPIYVHEVHNRTVVDGLKQGAVFIEEL 60
 QY 61 SDVPGSYLIFSAGVSEVQAEEROLTFPDATCPVTKVMQVAKAKQREYVILIG 120
 DB 61 DEVPDATVIFSAGVSEVQAEEROLTFPDATCPVTKVMQVAKAKQREYVILIG 120
 QY 121 HAHGPEVEGTMQYEKCTEGGGIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 DB 121 HAHGPEVEGTMQYSESGERGQIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 QY 121 HAHGPEVEGTMQYSESGERGQIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 DB 121 HAHGPEVEGTMQYSESGERGQIYLVTEPEDVANKVNPNDLAVYTQTLSTMTKVMV 180
 QY 181 DALRQPSIKQKQKDDICATONRODAVNDLAKISDLILVYSGSPSSNSNRRLRIAYOL 240
 DB 181 DALRQPSIKQKQKDDICATONRODAVNDLAKISDLILVYSGSPSSNSNRRLRIAYOL 240
 QY 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 DB 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 QY 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 DB 241 GKPAYLIDTYODLKQDMLEIEVGVGTAGASAPVYLVOEVIDLQKAMGETTSVRENSGI 300
 QY 301 EEKVVSIPKEIK 313
 DB 297 EGNMFEVPEIKR 309

RESULT 13
 ISPH_PASMO STANDARD; PRT; 314 AA.
 ID ISPH_PASMO STANDARD; PRT; 314 AA.
 AC P57960;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ISPH protein.
 GN ISPH OR LYTB OR PM1664.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;

DB 242 DAYWDMAGYLOPMBEGKXKGVTAAGASAPVLYREVLATITNGWHER--VREBGAER 299
 QY 303 KYVFSIPKELK 314
 DB 300 SIVFVLPKELR 311

RESULT 15
 ISPH_XYLFA STANDARD; PRT; 316 AA.

AC 092AS9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE isph protein.
 GN ISPH OR LYTB OR XP2416 OR PD1435.
 OS Xylella fastidiosa (Strain Temecula / ATCC 700964).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI TaxID=3371, 183190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=945C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Aencio M.,
 Alvares A.J.G., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Birones M.R.S.,
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carro D.M., Carrer H.,
 Colauto N.B., Colombo C., Costa P.F., Costa M.C.R., Costa-Dorcy C.M.,
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 Fraga J.S., Franca S.C., Franco M.C., Frome M., Furian L.R.,
 Gantier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 Ho P.L., Hohlreis J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Martins E.M.F., Matsukuma A.Y.,
 Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Niani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 de Paiva B.R., Pereira G.A.G., Pereira H.A. Jr., Peguero J.B.,
 Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 Zago M.A., Zatz M., Zeldanis J., Zerbini J.C., Zerbini A.L.,
 "The genome sequence of the plant pathogen Xylella fastidiosa,"
 Nature 406:151-159(2000).
 [2]
 R2 SEQUENCE FROM N.A.
 R3 STRAIN=temecula; ATCC 700964;
 RC MEDLINE=22421331; PubMed=12533478;
 RX Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
 Miyaki C.Y., Furian L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
 Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
 Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
 Carro D.M., de Oliveira R.C., Nunes L.R., Silveira W.J.,
 Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
 Martin C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
 Baia G.S., Blanco S.R., Brito M.S., Camargo F.S., Celestino A.V.,
 da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.P., Kishi L.T.,
 Leont S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
 de Souza A.A., Truffi D., Tsubako M.H., Tsukumo F., Yanai G.Y., Zago L.G.,
 Zerbini J.P., Simpson A.J.G., Almeida N.P. Jr., Zerbini J.C.,
 Kitajima J.P.;
 "Comparative analyses of the complete genome sequences of Pierce's
 disease and citrus variegated chlorosis strains of Xylella
 fastidiosa,"

RL J. Bacteriol. 185:1018-1026 (2003).
 CC -!- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
 CC (DMAPP) (By similarity).
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
 CC (last) step.
 CC -!- SIMILARITY: BELONGS TO THE ISPH FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB004050; AAF85215.1; -;
 DR EMBL: AB012558; AAC29279.1; -;
 DR F01: C82561; C82561.
 DR HAMAP: MF_00191; -; 1.
 DR InterPro: IPR003451; LYTB.
 DR Pfam: PF02401; LYTB; 1.
 DR TIGRFAMs: TIGR00216; isph LYTB; 1.
 DR Isoprene biosynthesis: Complete proteome.
 KW SEQUENCE 316 AA; 34704 MW; 1A2B80B9A98D334A CRC64;
 SQ

Query Match 59.6%; Score 967; DB 1; Length 316;
 Best Local Similarity 61.0%; Pred. No. 1.7e-61;
 Matches 191; Conservative 53; Mismatches 67; Indels 2; Gaps 1;

1 MOYLANRPGGAGDRAIRIYDOLIEAFGAPTYRHHVNRRTVVDGKOGAVFTEL 60
 1 MYVLANRPGGAGDRAIRIYDOLIEAFGAPTYRHHVNRRTVVDGKOGAVFTEL 60
 61 SEVPAGSYLIFSAHSGVSEVQEAEROLTVFATCPVTKVMQVAKAKGAREVILG 120
 61 HGVPGATVIFSAHSGVSAVROAQRGKVPDACPVTXKHLVABHRCGRDMLIG 120
 121 HAGEPEVBCMGQYKCTBGGGIYVETPEYRNLKANNNDLAVVOTLTSMOTKAV 180
 121 HAGEPEVBCMGQYKCTBGGGIYVETPEYRNLKANNNDLAVVOTLTSMOTKAV 180
 121 HAGEPEVBCMGQYKCTBGGGIYVETPEYRNLKANNNDLAVVOTLTSMOTKAV 180
 181 DALRPOPSIKKQKDDICVATONRQDAVHDIAKISDILLVGSFNSNSNLRIRIAVL 240
 181 DALRPOPSIKKQKDDICVATONRQDAVHDIAKISDILLVGSFNSNSNLRIRIAVL 240
 241 GRPVLIDTYYDLPKQDMLGIEVGVTAAGASAPVLYREVLATITNGWHER 300
 241 GVASYLIDTYYDLPKQDMLGIEVGVTAAGASAPVLYREVLATITNGWHER 300
 301 EKVFSPKELK 313
 299 PESWVFLPKELR 311

Search completed: January 29, 2004, 15:50:43
 Job time : 6.98823 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 Search time 23.7034 Seconds
(without alignments)
3461.975 Million cell updates/sec

Title: US-09-941-947a-18
Perfect score: 1623
Sequence: 1 MGVLANPRGFCAGVDRAIE.....GIEKVVPSIPKELKGMQA 318

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTRMEL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	989	60.9	320	16	Q8DET0
2	958	59.0	318	16	Q8EB17
3	866	53.4	315	16	Q8D2R2
4	764	47.1	336	16	Q8FOP0
5	751	46.3	372	16	Q8G257
6	658	40.5	352	16	Q8G418
7	628	38.7	330	16	Q8F313
8	599	36.9	535	5	Q8I295
9	598.5	36.9	535	5	Q8BUX6
10	404.5	24.9	827	16	Q8R152
11	393	24.2	299	16	Q8EMR9
12	358	22.1	642	16	Q87109
13	343	21.1	402	16	Q8DX29
14	305	18.8	464	10	Q8FEP0
15	300	18.5	466	10	Q84B35
16	265	16.3	332	16	Q8KFN9

17	230	14.2	452	10	Q9SZ00	Q9SZ00 arabidopsis
18	176.5	10.9	167	10	Q9SP19	Q9SP19 nicotiana t
19	170	10.5	221	10	Q94EP5	Q94EP5 brassica ra
20	170	10.5	223	10	Q94ER4	Q94ER4 brassica ra
21	150.5	9.3	130	10	Q8LKH3	Q8LKH3 forsythia l
22	115	7.1	929	16	Q86810	Q86810 streptococ
23	112.5	6.9	434	16	Q8S5X9	Q8S5X9 thermotoga
24	112	6.9	969	16	Q8G4H3	Q8G4H3 bifidobacte
25	111.5	6.9	670	2	P96626	P96626 bacillus su
26	111.5	6.9	777	16	Q31489	Q31489 bacillus su
27	110	6.8	500	16	Q8YA67	Q8YA67 listeria mo
28	110	6.8	542	2	Q8RR20	Q8RR20 listeria mo
29	109	6.7	499	16	Q92EY8	Q92EY8 listeria in
30	109	6.7	925	5	Q22758	Q22758 caenorhabdi
31	108.5	6.7	596	16	Q97MA7	Q97MA7 clostridium
32	108.5	6.7	602	16	Q8RBX6	Q8RBX6 thermomater
33	108	6.7	886	16	Q05560	Q05560 mycobacteri
34	107.5	6.6	501	5	Q61118	Q61118 leishmania
35	105.5	6.5	960	11	Q8BSZ1	Q8BSZ1 mus musculu
36	104.5	6.4	527	16	Q98AG2	Q98AG2 rhizobium l
37	104	6.4	488	16	Q8CMQ7	Q8CMQ7 staphylococ
38	103.5	6.4	482	15	Q99F12	Q99F12 chimpanzee
39	103.5	6.4	545	16	Q8CX13	Q8CX13 oceanobacill
40	103.5	6.4	960	11	Q8C1E4	Q8C1E4 mus musculu
41	103	6.3	374	16	Q8X8F4	Q8X8F4 escherichia
42	103	6.3	582	5	Q9U5L7	Q9U5L7 paracentrot
43	103	6.3	713	16	Q8U9Z5	Q8U9Z5 agrobacteri
44	102.5	6.3	349	1	Q49598	Q49598 methanopyru
45	102.5	6.3	396	17	Q96Y01	Q96Y01 sulfobobus

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	320 AA.
Q8DET0			
AC Q8DET0			
DT 01-MAR-2003 (TREMBLrel. 23, Created)			
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DN Penicillin tolerance protein.			
OS V10504.			
OS Vibrio vulnificus.			
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;			
OC Vibrionaceae; Vibrio.			
OX NCBI_TaxID=672;			
RN [1]			
RP SOURCE FROM N.A.			
RA STRAIN=CMCP6;			
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,			
RA Choy H.E.;			
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";			
RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.			
DR EMBL; AE016798; AAO09023.1; ..			
KW Complete proteome.			
SQ SEQUENCE 320 AA; 35575 MW; 097B45ACDA1B1D8 CRC64;			

QY	1	MOVLANPRGFCAGVDRAIEIVDALEAFCAPIYVRRHVRRTVVDLKGAVFIETL	60
QY	1		60
QY	61	SDVPVGSYLIFSAHGVSKVQDAEERQITVDATCPITVYVMQVAKAKQGREVILIG	120
QY	61		120
QY	62	HEVPDNIYIFSAHGVSKVQDAEERQITVDATCPITVYVMQVAKAKQGREVILIG	124
QY	62		124
QY	121	HAGHVEGTGQVYKCTGGGIVYVETPVEDVRNK--VNNNDLAVYQTTLSMTDFKV	178
QY	121		178
QY	122	HAGHVEGTGQVYKCTGGGIVYVETPVEDVRNK--VNNNDLAVYQTTLSMTDFKV	182
QY	122		182

QY 179 MYDALREOPSIKEOKKDDICATATONRODAVHDLAKISDLILVYGSPPSSNSNRLRELAIV 238
 DB 183 VIDELARRVPELIGRKDDICATATONRODAVRELAKDDVYVYVGSKSSNSTLKEIAE 242
 QY 239 QLGKPAVILDTYODIKQDMLGIEVGVYTAGASAEVLVQVVDQLKAMGGETTSVRENS 298
 DB 243 KLGTFAVILDTYODIEPQDFGKAVGVYTAGASAEVLVQVVDQLKAMGGETTSVRENS 300
 QY 299 GIEKVPFISIPKELK 313
 DB 301 GREENMFPEVPELQ 315

RESULT 2

Q8EB17 PRELIMINARY; PRT; 318 AA.
 ID Q8EB17
 AC Q8EB17
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Penicillin tolerance protein LytB.
 GN LYTB OR S03529.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 CX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12366813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eisen J.A., Seshadri R., Ward N., Melhe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
 Debey R.T., Pederson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Imbra M., Lee K., Berry K., Lee C.,
 Mueller J., Klotzel H., Gill J., Ueberbach T.R., McDonald L.A.,
 Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123 (2002).
 DR EMBL; AF015789; AAN56520.1; -.
 DR TIGR; S03529; -.
 KW Complete proteome.
 SQ SEQUENCE 318 AA; ADCE10B6A3582E82 CRC64;

Query Match 59.0%; Score 958; DB 16; Length 318;
 Best Local Similarity 60.1%; Pred. No. 5.7e-64;
 Matches 188; Conservative 50; Mismatches 71; Indels 4; Gaps 2;

QY 1 MQVLANPRGFCAGVDRAIEIVDAIEAFGAPIYVREHVHNRVYVDGLKQKGAFFIEL 60
 DB 10 LNIKLANPRGFCAGVDRAISIVERALELFSPEIYVREHVHNRVYVDGLKQKGAFFIEL 69
 QY 61 SDVPVGSYLIFSAHGVSKVQOAEEROLTVFATCPVTVKVMQVAKHAKQGEVLIG 120
 DB 70 DQVFNNTVIFSAHGVSKVQOAEEROLTVFATCPVTVKVMQVAKHAKQGEVLIG 129
 QY 121 HAHGEVEGTMGQYEKTEGGGIYLVETPEDEVNLIKVNPNDLAVYVOTLLSMTDTKMV 180
 DB 130 HAHGEVEGTMGQYD--NPNNGSVYLISPADVETLEVDPNNLCTVOTLLSVDDTLDI 187
 QY 181 DALREOPPSIKEOKKDDICATATONRODAVHDLAKISDLILVYGSPPSSNSNRLRELAIV 240
 DB 188 SALKRFPISIEGRKDDICATATONRODAVHDLAKISDLILVYGSPPSSNSNRLRELAIV 247
 QY 241 GKPAVILDTYODIKQDMLGIEVGVYTAGASAEVLVQVVDQLKAMGGETTSVRENSGI 300
 DB 248 GTOSTVLTADDDISWENITKAVTVAGASAEVLVQVVDQLKAMGGETTSVRENSGR 305
 QY 301 BEKVPFISIPKELK 313

DB 306 KEDTVFAVPAELR 318

RESULT 3

Q8D2R2 PRELIMINARY; PRT; 315 AA.
 ID Q8D2R2
 AC Q8D2R2
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE LytB protein.
 GN LYTB.
 OS Myxlesworthia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Myxlesworthia.
 CX NCBI_TaxID=164609;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 Aksoy S.,
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 flies, Myxlesworthia glosiniida.";
 RL Nat. Genet. 32:402-407 (2002).
 DR EMBL; AB063521; BAC24438.1; -.
 KW Complete proteome.
 SQ SEQUENCE 315 AA; 6552098712A2884 CRC64;

Query Match 53.4%; Score 866; DB 16; Length 315;
 Best Local Similarity 51.9%; Pred. No. 4.6e-57;
 Matches 162; Conservative 72; Mismatches 76; Indels 2; Gaps 1;

QY 1 MQVLANPRGFCAGVDRAIEIVDAIEAFGAPIYVREHVHNRVYVDGLKQKGAFFIEL 60
 DB 1 MNIKLANPRGFCAGVDRAIIVRAIEIYGPPIYVREHVHNRVYVDGLKQKGAFFIEL 60
 QY 61 SDVPVGSYLIFSAHGVSKVQOAEEROLTVFATCPVTVKVMQVAKHAKQGEVLIG 120
 DB 61 CDVPSKILIFSAHGVSKVQOAEEROLTVFATCPVTVKVMQVAKHAKQGEVLIG 120
 QY 121 HAHGEVEGTMGQYEKTEGGGIYLVETPEDEVNLIKVNPNDLAVYVOTLLSMTDTKMV 180
 DB 121 HXNHPHIGTIGQYNNPNK--KVFVIGSTIEICLAKIKDNNLFFYOTLLSVDDTKII 178
 QY 181 DALREOPPSIKEOKKDDICATATONRODAVHDLAKISDLILVYGSPPSSNSNRLRELAIV 240
 DB 179 FAIKKKFPIIEPRKDDICATATONRODAVHDLAKISDLILVYGSPPSSNSNRLRELAIV 238
 QY 241 GKPAVILDTYODIKQDMLGIEVGVYTAGASAEVLVQVVDQLKAMGGETTSVRENSGI 300
 DB 239 GKSYLIDTYKEIKKSLNGLVNNIGITKASAEVLVQVVDQLKAMGGETTSVRENSGI 298
 QY 301 BEKVPFISIPKEL 312
 DB 299 IEKTFMIPKEL 310

RESULT 4

Q8F0P0 PRELIMINARY; PRT; 336 AA.
 ID Q8F0P0
 AC Q8F0P0
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative isoprenyl diphosphate biosynthesis protein Isph.
 GN CE1079.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 CX NCBI_TaxID=152794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y8-314 / AJ 12310 / DSM 44549 / JCM 11189;

RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.,
 RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314."
 DR Submitted (May-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AP005217, BAC17889.1; -.
 KW Complete proteome.
 SQ SEQUENCE 336 AA; 36430 MW; C47B456FEEB82AC7 CRC64;

Query Match 47.1%; Score 764; DB 16; Length 336;
 Best Local Similarity 48.4%; Pred. No. 2,3e-49;
 Matches 153; Conservative 68; Mismatches 83; Indels 12; Gaps 5;

QY 2 QIVLANPFGCAGVDRAIEIVDQIAEAFGAPIVYREHVNRITVVDGLKQKQAVFIEIS 61
 DB 26 KILIAAPRGYAGVAVTVERALVEEYGAPIYVREIYVNRVYVDLTAEKGVIFVEAS 85
 QY 62 DVPVSYLIFSAHGVSKVQGEAREOLTFPDATCPVTKYHQAQAKQAREVILIGH 121
 DB 86 EAPBEGAHVFSAGVSPAVKEPAKALQADACPLVTKVNEVRRPDQKQFILLFIGH 145
 QY 122 AGHPEVEGTMGQYEKTEGGGIYVETPEDEVRLK---VNNENDLAYVOTTLSTMTDKV 178
 DB 146 EGHPEVEGTMGHSLDRT-----HIVDGIIESIPGLAFADBN-LIYLSQTLISVDTEME 199
 QY 179 MVDALREQPSIKQKDDICATQNRQDAVHDLAKISDLILVSGSPNSNSNRILEIAV 238
 DB 200 IYRELKRVYPLQDPPSDICATQNRQDAVAVAKVIERCDLMIVGTSNSNSVRILVEAL 259
 QY 239 QUG-KPAYLIDTYQDLKODMLEGIEVGTAGASPEVLYOGVIELKAMGETTSVSEN 297
 DB 260 QAGAAHVLIVYAHQIDEMALDVGCTIGISSASPELITVGLERLAGYGH--DVSEV 317
 QY 298 SGIEEKVVFSPKELK 313
 DB 318 TTAERKIVFALPRVLR 333

RESULT 5

ID Q8G257 PRELIMINARY; PRT; 372 AA.
 AC Q8G257;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Penicillin tolerance protein lypB.
 DE lypB OR BR0475.
 OS Brucella suis.
 GN Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RC MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beaman M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson M.C., Ayodeji B., Kraul M., Shetty J., Malek U., Van Aken S.E.,
 RA Rietmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.B., Hailing S.M., Boyle S.M., Fraser C.M.,
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL: AE014357; AAN29418.1; -.
 DR TIGR: BR0475; -.
 KW Complete proteome.
 SQ SEQUENCE 372 AA; 40930 MW; 365B9906518CC9BC CRC64;

Query Match 46.3%; Score 751; DB 16; Length 372;
 Best Local Similarity 47.7%; Pred. No. 2.6e-48;
 Matches 154; Conservative 59; Mismatches 80; Indels 22; Gaps 5;

QY 1 MQIVLANPFGCAGVDRAIEIVDQIAEAFGAPIVYREHVNRITVVDGLKQKQAVFIEEL 60
 DB 34 LEIRLCGRGRCAGVDRAIQIVLALCKKGAIPYVRHIVNRVYVDEGLQARGAIFVEAL 93
 QY 61 SDVPV---GSTYLFSAHGVSKVQGEAREOLTFPDATCPVTKYHQAQAKQAREVILIGH 117
 DB 94 DRIPAHNRQPVFSAGVSPAVKEPAKALQADACPLVTKVNEVRRPDQKQFILLFIGH 153
 QY 118 LIGHAGPEVEGTMGQYEKTEGGGIYVETPEDEVRLK---VNNENDLAYVOTTLSTMTDKV 177
 DB 154 LIGHAGPEVEGTMGQ-----LPDQAVLILVEEDAHCHPDEENLIGVYOTTLISVDTE 209
 QY 178 MVDALREQPSIKQKDDICATQNRQDAVHDLAKISDLILVSGSPNSNSNRILEIAV 237
 DB 210 GIIEKIQARFPAALPAFESICATQNRQDAVRAAPCGDFLIVGAPNSNSNRILEIAV 269
 QY 238 VQEG-KPAYLIDTYQDLKODMLEGIEVGTAGASPEVLYOGVIELKAMGETTSVSEN 290
 DB 270 EKQGRFMSMLVQRAEDIEWEGIDISVGLSAGASAPRILVDEILIDAFKAPDVYKIELAE 329
 QY 291 TTSVRENSGIEEKVVFSPKELK 313
 DB 330 TTVETEN-----FLVNRER 344

RESULT 6

ID Q8G418 PRELIMINARY; PRT; 352 AA.
 AC Q8G418;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE lspH protein involved in conversion of
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate into IPP and
 DE DMAP.
 DE lspH OR BL1361.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=216816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCC 2705;
 RC MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karimianzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
 RA Pidmore R.D., Arigoni F.,
 RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation
 RT to the human gastrointestinal tract."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL: AE014766; AAN25161.1; -.
 DR Complete proteome.
 SQ SEQUENCE 352 AA; 37863 MW; 984B6BF80286D136 CRC64;

Query Match 40.5%; Score 658; DB 16; Length 352;
 Best Local Similarity 43.8%; Pred. No. 2.3e-41;
 Matches 145; Conservative 64; Mismatches 96; Indels 26; Gaps 7;

QY 3 VIANPFGCAGVDRAIEIVDQIAEAFG-----APIYREHVNRITVVDGLKQ 51
 DB 27 VVADPFGCAGVDRAIVQTLIKAAASGKRTREDDLPVYVRQIVNRKVVVEDIAG 86
 QY 52 KGAVFIEELSDVP-----VGSYLFSAHGVSKVQGEAREOLTFPDATCPVTKYHQAQAKQAREVILIGH 106
 DB 87 QGAVFVQELAEIPDAQAAGIPVFSAGVSPAVKEPAKALQADACPLVTKVNEVRRPDQKQFILLFIGH 146
 QY 107 AKIAQGREVILLIGHAGPEVEGTMGQYEKTEGGGIYVETPEDEVRLK---VNNENDLAYV 166
 DB 147 LRFVRSRGYRIVYIGHKGEAVGVSGSP-----HYHLIEBSDDVDSLPADTKLVLL 201
 QY 167 TQTLISMTDKVAVDALREQPSIKQKDDICATQNRQDAVHDLAKISDLILVSGSPNSNRILEIAV 226
 DB 202 SQTTLISVDTEADTIALKAKFPWIOEPSSDICATQNRQDAVAVAKVIERCDLMIVGTSNSNSVRILVEAL 261

QY 227 SSSNRRLREIAVO-LGK--PAYLIDTYODLTKODMLEGIEVGVTAAGASAEVLYVOEVIDQ 283
 DB 262 SSSNRRLREIAVO-LGK--PAYLIDTYODLTKODMLEGIEVGVTAAGASAEVLYVOEVIDQ 321
 QY 284 LKAWGERTSVRENSGIEKRVFSPKELK 314
 DB 322 LONLIG--FTGMRKSVETIKENMHFVLPALRLR 350

RESULT 7

Q88313 PRELIMINARY; PRT; 330 AA.

AC Q88313; 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Penicillin tolerance protein.
 GN LYTB OR LA2420.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 OX NCBI_TaxID=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;
 RA Ren S.
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE011409; AAK49619.1;
 KW Complete proteome.
 SQ SEQUENCE 330 AA; 36940 MW; 5D7642CD6AE3F86E CRC64;

Query Match 38.7%; Score 628; DB 16; Length 330;
 Best Local Similarity 47.1%; Pred. No. 3.8e-39;
 Matches 136; Conservative 57; Mismatches 82; Indels 14; Gaps 7;

QY 2 QVLANPRGCGVDRALIEYDQALFAFGAPLYVREHVNRTVVDGKQKGAFTIEBLS 61
 DB 22 KLYLANPRGCGVDRALIEYDQALFAFGAPLYVREHVNRTVVDGKQKGAFTIEBLS 81
 QY 62 DVPVGSYLIFSAHGVSKVQOEAEROLTVFDATCPLYTKVMQVAKAKQREYILIGH 121
 DB 82 EAPNAGATVIFSAHGVSPVVEAKQKGMIGATCPLYTVRHRK-ARKIKDTHQIYIGH 140
 QY 122 AGHPVEVGMQYKCTBGGGIYVETPEDVRLKYN-NPN-DLAVTQTTLSMTDTKYM 179
 DB 141 EGDDEAIGTMBEAE-----MFLVESLDDIISLKDIPNKRLTYLQTTLSVADTKNI 193
 QY 180 VDALRQFPISIKQKQKODICATONRODAVHDLAKISDLILVGVSPSSNSNRLEIAVO 239
 DB 194 IDQISTFFVEHPSKDDICATTERQEAVALMKDIDMLVIGADNSNSRLQLA-Q 252
 QY 240 LGKP-AYLIDTYODLTKODMLEG-IEVGVTAAGASAEVLYVOEVIDQK 285
 DB 253 KSKPESHFKVSTADLSKEYIYONNEIKILGLITAGASTPYQVLYVDEITSKLK 301

RESULT 8

Q81295 PRELIMINARY; PRT; 535 AA.

AC Q81295; 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE LYTB protein.
 GN LYTB
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36339;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22255708; PubMed=12368867;
 RA Hall N., Pain A., Bettman M., Churcher C., Harris B., Harris D.,
 Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,

RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Cotton C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornum A.,
 RA Knights A., Kontoforov B., Kyes K., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Tulston J.E., Craig A., Newbold C., Barrett B.G.
 RT "Sequence of plasmodium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 DR EMBL; AL011745; CAD49005.1;
 SQ SEQUENCE 535 AA; 62470 MW; E7CB6FCCD22FA07E CRC64;

Query Match 36.9%; Score 599; DB 5; Length 535;
 Best Local Similarity 41.4%; Pred. No. 1.1e-36;
 Matches 130; Conservative 58; Mismatches 120; Indels 6; Gaps 2;

QY 3 IVLANPRGCGVDRALIEYDQALFAFGAPLYVREHVNRTVVDGKQKGAFTIEBLS 62
 DB 222 LYLINPRGCGVDRALIEYDQALFAFGAPLYVREHVNRTVVDGKQKGAFTIEBLS 281
 QY 63 VPVGSYLIFSAHGVSKVQOEAEROLTVFDATCPLYTKVMQVAKAKQREYILIGH 122
 DB 282 VPDHILILYSAHGISPQREIAKKKQKLEIDATCPLYTKVMQVAKAKQREYILIGH 341
 QY 123 GHPVEVGMQYKCTBGGGIYVETPEDVRLKYN-NPN-DLAVTQTTLSMTDTKYM 182
 DB 342 NHVEVIGTMBEAE-----HIVENVDVDMKPNPKRLFYVQTTLSMDALYQK 396
 QY 183 LRQFPISIKQKQKODICATONRODAVHDLAKISDLILVGVSPSSNSNRLEIAVO 242
 DB 397 LKAFPHETIPSSICATINROTALNKICTKDDIITVGVSSSSNAKKLVYSQIRNV 456
 QY 243 PAYLIDTYODLTKODMLEGIEVGVTAAGASAEVLYVOEVIDQKMGERTSVRENSGIEE 302
 DB 457 PAVLINTVHDDQOILKVNKIALTSAASTPEQTKQFVNLTMPFNVTYQNTDGAHEN 516
 QY 303 KVPSPKELKGM 316
 DB 517 VPKMKLPNML-LHM 529

RESULT 9

Q9BX6 PRELIMINARY; PRT; 535 AA.

AC Q9BX6; 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE LYTB.
 GN LYTB
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21172655; PubMed=11274098;
 RA Alincic B., Kollas A.K., Sanderbrand S., Wiesner J., Hintz M.,
 RA Beck E., Juma H.,
 RT "Gp6 is involved in the 2-C-methyl-D-Erythritol 4-Phosphate pathway
 of isoprenoid biosynthesis in Escherichia coli."
 RL J. Bacteriol. 183:2411-2416(2001).
 DR EMBL; AF333927; AAK12102.1;
 DR InterPro; IPR003451; LYTB.
 DR InterPro; IPR002052; N6_Mtase.
 DR Pfam; PF02401; LYTB; 1.
 DR TIGRFAMs; TIGR00216; Isph_LYTB; 1.
 DR PROSITE; PS00092; N6_MTASE; 1.

SQ SEQUENCE 535 AA; 62504 MW; E7C6CDDC22FA07E CRC64;
 Query Match 36.9%; Score 598.5; DB 5; Length 535;
 Best Local Similarity 43.6%; Pred. No. 1,3e-36;
 Matches 123; Conservative 55; Mismatches 99; Indels 5; Gaps 1;
 QY 3 IVIANRGFCAGVDAIRIEIVDAIEAFGAPIYVREHVNRKTVVDGLKQKGAFFIEBLSD 62
 DB 222 LVLINRGFCAGVDAIRIEIVDAIEAFGAPIYVREHVNRKTVVDGLKQKGAFFIEBLSD 281
 QY 63 VAVGSVTLISAGVSGVNOQAEERQTLTPDAPLTVTVNQVAKHAKQSEVILLIGA 122
 DB 282 VDGHTLLYSAGISQIRIENAKKKLISIDTCPLVNVYHYVQKAKENDIILLIGK 341
 QY 123 GHPEVEGTGQYEKTEGGGIVLVEPVEDVRLKVNPNDAVYVQTLSMTPTKVMYA 182
 DB 342 NVEVEVIGTNEAPHQI-----HIVEVNDVVDKLNPLNKKLPYVQTLTSMDCALIVQK 396
 QY 183 IREQPSIKEQKKDICIATONRODAVHDLAKISDILLVSGSPNSNRLREIVQLOK 242
 DB 397 LKNKPEHIEIIPSGSICVATINRQTLNKKICTKCDLTIIVGSSSSNAKKLVYSSQIRVY 456
 QY 243 PAVLIDTVQDLKQDMLEGIEVVGTVTAGASAPVLYQVEVIDL 284
 DB 457 PAVLIDTVQDLKQDMLEGIEVVGTVTAGASAPVLYQVEVIDL 498
 RESULT 10
 OBR152 PRELIMINARY; PRT; 827 AA.
 AC OBR152;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE LysB protein.
 GN FN1781.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
 CC Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapetral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grechlin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
 RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
 RA Fongstein M., Kyriades N., Overbeek R.;
 RT Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL, AE010480; AAL93880.1; -
 DR InterPro; IPR003451; LysB.
 DR InterPro; IPR003029; SL.
 DR Pfam; PF02401; LysB; 1.
 DR Pfam; PF00575; SL; 4.
 DR SMART; SM00316; SL; 6.
 DR TIGRfams; TIGR00216; 1spH_LysB; 1.
 DR PROSITE; PS0126; SL; 4.
 DR Complete proteome.
 SQ SEQUENCE 827 AA; 94507 MW; 23871DCEDD082A35E CRC64;

Query Match 24.9%; Score 404.5; DB 16; Length 827;
 Best Local Similarity 30.7%; Pred. No. 8.9e-22;
 Matches 103; Conservative 68; Mismatches 119; Indels 45; Gaps 9;
 QY 1 MOIVIANRGFCAGVDAIRIEIVDAIEAFGAPIYVREHVNRKTVVDGLKQKGAFFIEB 59
 DB 1 MOIVIANRGFCAGVDAIRIEIVDAIEAFGAPIYVREHVNRKTVVDGLKQKGAFFIEB 59
 QY 60 -----LSDDVPGSVYLISAGVSGVNOQAEERQTLTPDAPLTVTVNQVAKHAKQSEV 114
 DB 60 -----LSDDVPGSVYLISAGVSGVNOQAEERQTLTPDAPLTVTVNQVAKHAKQSEV 114

DB 60 ELLNDMDLEKEDDIIVIRAHGTSKSVHEKLEKRVKVPDATICFVNKIRQELIANENGY 119
 QY 115 EVILLIGHGHEVGTGQYEKTEGGGIVLVEPVEDVRLKVNPNDAVYVQTLSMT 174
 DB 120 SILFPGDNHPEVGVV-----SFADIDQIFESFEKAKKIDIDKTYLLSTQTLN-- 171
 QY 175 DTKVVDALREQPSIKEQK-----DDICATONRODAVHDLAKISDILLVSGSP 225
 DB 172 -----KKFBEIKKYFENKVNVIIPKICATVNRQKAVEDLVAKVEVMIIVDT 222
 QY 226 MSSNSNRLREIVQLOKGPAYLIDTVQDLKQDMLEGIEVVGTVTAGASAPVLYQVEIDOLK 285
 DB 223 KSNNTKLYEISKTLNDSYLVNEBQDLISIPKGEVVGITAGASTPEETIMNI--SKK 280
 QY 286 AMG-GETTSVRENGS-----TEEKVPSIPKELKK 314
 DB 281 VRIYKMSNVNENQERSLMEB-----FLPNQEKR 311

RESULT 11
 OBR152 PRELIMINARY; PRT; 299 AA.
 AC OBR152;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE LysB protein.
 GN MYPEL1330.
 OS Mycoplasma penetrans.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HF-2;
 RX MEDLINE=22354719; PubMed=12466555;
 RA Saeki Y., Ishikawa J., Yamashita A., Oshina K., Kent T., Furuya K.,
 RA Yoshino C., Horino A., Shida T., Saeki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans."
 RL Nucleic Acids Res. 30:5293-5300(2002).
 DR EMBL; AP004170; BAC43925.1; -
 DR Complete proteome.
 SQ SEQUENCE 299 AA; 34190 MW; CFF14G65FB55F9E CRC64;

Query Match 24.2%; Score 393; DB 16; Length 299;
 Best Local Similarity 30.2%; Pred. No. 1.5e-21;
 Matches 93; Conservative 68; Mismatches 99; Indels 48; Gaps 8;

QY 1 MOIVIANRGFCAGVDAIRIEIVDAIEAFG-----APIYVREHVNRKTVVDGL 49
 DB 1 MOIVIANRGFCAGVDAIRIEIVDAIEAFG-----APIYVREHVNRKTVVDGL 49
 QY 50 KQKQ-----AVFTELSDVPGSVYLISAGVSGVNOQAEERQTLTPD 93
 DB 51 EELGIGTQDMKNSRSEIISIEIKDKNNPFI--VIFSAGTQKTDKAREKGLVFD 107
 QY 94 ATCPYLVYVQVAKHAKQSEVILLIGHAGPEVGTGQYEKTEGGGIVLVEPVEDV 153
 DB 108 TTCIYVPTKDLIRKEKEQGYQIFYIGVNNHPEITISL-SIDK-----SILLIEVNDIE 161
 QY 154 NLKVNPNDAVYVQTLSMTDKVMDALREQPSIKEQKKDICIATONRODAVHDL 213
 DB 162 NRTESKEPIVYVQTLSMTISIEFEIETELSKRNI--EIKDICNAKRODAVINMP 219
 QY 214 KISDILLVSGSPNSNRLREIVQLOKGPAYLIDTVQDLKQDMLEGIEVVGTVTAGASAP 273
 DB 220 SEVDLLVVGDIKNSNKKLVIGIKQIBSHLIMTGNRIDEFINKKCLATISGCTP 279
 QY 274 EVLYQVEVI 281
 DB 280 TWLANVYI 287

RESULT 12
 097109 PRELIMINARY: PRT; 642 AA.
 AC 097109: 18, Created)
 DT 01-OCT-2001 (TREMELREL. 18, Last sequence update)
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)
 DE Fusion penicillin tolerance lytB domain (N-term) and S1 ribosomal protein (C-term).
 GN CAC1847.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Bleton G., Ometchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.W., Dubois J., Ohi D., Hilt J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucet-Stamm L., Soucaille P., Daly M.D.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AS007693; AAK79811.1; -
 DR InterPro; IPR003451; LytB.
 DR InterPro; IPR000110; Ribosomal_S1.
 DR InterPro; IPR003029; S1.
 DR Pfam; PF02401; LytB; 1.
 DR Pfam; PF05575; S1; 3.
 DR PRINTS; PR00681; RIBOSOMAL_S1.
 DR SMART; SM00316; S1; 4.
 DR TIGRFAMs; TIGR00216; lspb_lytB; 1.
 DR PROSITE; PS0126; S1; 3.
 DR Complete proteome.
 SQ SEQUENCE 642 AA; 71968 MW; A8833AA24420EDB CRC64;
 Query Match 22.1%; Score 358; DB 16; Length 642;
 Best Local Similarity 27.1%; Pred. No. 1.9e-18;
 Matches 90; Conservative 77; Mismatches 129; Indels 36; Gaps 8;
 QY 2 OVLANPGRGFCAGVDRAIEIVDQATEAFGAP-IVYRHEVNRRTVVDGKQKGA--VFTEE 59
 DB 3 KWLAEKAFGFCGVRADVALLTQEKYKLIYTLGELHNDVVDKLDNNVPTIGIED 62
 QY 60 LSDVPVGSYLIFSAHGVSKVOEAEERQITVFDAITCPVLTAKVHQAQKQREVLII 119
 DB 63 IMLKENDVTLIRSHGISSEIYKILSKGLVIANCTPPTKIQEKVKKYMLGVDIV 122
 QY 120 GHAGHPVEVGIMQYBKECTEGGGLVETPEDEVRLKNNPNDLAVYTQTLSTMDTKM 179
 DB 123 GKHYPEVINGIMCD-----NKAIISKQENLLEN--ITSEKVCIVSQTEKLENMERY 175
 QY 180 VVALBEOPFISKEOKKDDICVATONRODAVHDLAKISDILLVAGSPNSNSRLAEIAYO 239
 DB 176 LKEVKNR--AIVISFNTICATISEROKIADLSKRDPMVITGKQSNITKLYEICKS 233
 QY 240 LKRPAYLIDTYODLKODMLEGIE-VGVYTAGASAPVLEVOEVIDQL--KAMGETTS--- 293
 DB 234 NQNETIHWENSEGIEPENILKNNKQICIGVYAGASTPDWIIIEAISQSEINQSTNETNEMA 293
 QY 294 -----VRENSG-----IEEKVVF 306
 DB 294 DAMKPIAHEGKIYGAASYTGELIIOVSSEKVF 325
 RESULT 13
 08DK29 PRELIMINARY: PRT; 402 AA.
 AC 08DK29: 23, Created)
 DT 01-MAR-2003 (TREMELREL. 23, Last sequence update)

DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)
 DE T11041 protein.
 GN T11041.
 OS Synecchococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=BP-1;
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1.
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005372; BAC08594.1; -
 DR Complete proteome.
 SQ SEQUENCE 402 AA; 45213 MW; FB0851B618B73510 CRC64;
 Query Match 21.1%; Score 343; DB 16; Length 402;
 Best Local Similarity 30.7%; Pred. No. 1.3e-17;
 Matches 109; Conservative 53; Mismatches 105; Indels 88; Gaps 14;
 QY 3 IVLANPGRGFCAGVDRAIEIVDQATEAFGAP-IVYRHEVNRRTVVDGKQKGA--VFTEE 57
 DB 57 IRLAEAFGFCGVRADVALLTQEKYKLIYTLGELHNDVVDKLDNNVPTIGIED 116
 QY 58 --BELSDVPVGSYLIFSAHGVSKVOEAEERQITVFDAITCPVLTAKVHQAQKQREVLII 115
 DB 117 GVXDFSGVRPDVVLPAFGASVQEMQLNERGCTIVDTTCVWVSKWSEYKHKVSFT 176
 QY 116 VILGHAGHPVEVGIMQYBKECTEGGGLVETPEDEVRLKNNPNDLAVYTQTLSTMDTKM 155
 DB 177 SLHGKXNHEHTIT-----SSPAGTYLL-----VLANLEAVYVDYLLHGDRAPFM 224
 QY 156 ---KVNPE-----NDLAVY--TQTLSTMDT-----KVN-----DALBEQPSI 190
 DB 225 AKPAKASPGFDPDRDVRVQIANQTMKQETFOIKLPEFTIMIOKXGPRLEHFWSP 284
 QY 191 KEOKKDDICVATONRODAVHDLAKISDILLVAGSPNSNSRLAEIAYOQKPAVLI 249
 DB 285 -----NTICATQERQDAMSLVKEPLDKVYVIGSYNSNTTHQELAIHIGITSYRIDS 339
 QY 250 -----YOD--LKQDMLEGIEV-VGVYTAGASAPVLEVOEVIDQLKA 286
 DB 340 ADRIGPGRIRHKEPLDQNPVAAEWLDPDRPITIGITSGASTPDKVEEVLAKIPA 394
 RESULT 14
 09FEPO PRELIMINARY: PRT; 464 AA.
 AC 09FEPO: 16, Created)
 DT 01-MAR-2001 (TREMELREL. 16, Last sequence update)
 DT 01-MAR-2001 (TREMELREL. 16, Last annotation update)
 DE LytB-like protein precursor.
 GN LytB.
 OS Adonis palaeatina.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
 OC Ranunculaceae; Adonis.
 OX NCBI_TaxID=151078;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Immature and developing flower buds;
 RX MEDLINE=20461233; PubMed=11004185;
 RA Cunningham P.X., Jr., Lafond T.P., Gantt B.;
 RT "Evidence of a Role for LytB in the Nonmevalonate Pathway of
 RT Isoprenoid Biosynthesis."
 RL J. Bacteriol. 182:5841-5848(2000).
 DR EMBL; AF270978; AAG21984.1; -

